ECONOMICS

IN TERMS OF
THE GOOD, THE BAD
AND THE ECONOMIST

Matt McGee
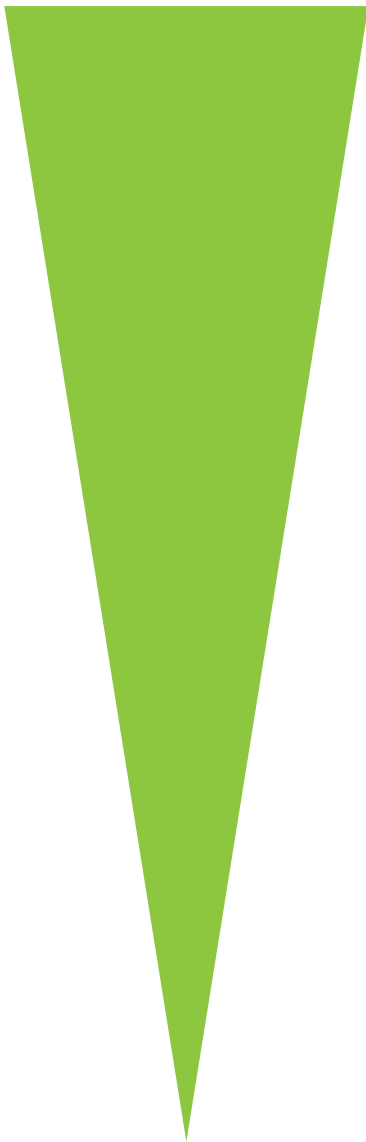
3rd Edition

ECONOMICS

IN TERMS OF
THE GOOD, THE BAD
AND THE ECONOMIST

Matt McGee

3rd Edition

First published in 2012 by IBID Press, Australia.

Library Catalogue:

1. Economics

2. International Baccalaureate

3. Matthew McGee

Cover design by Crystallisation.

Associated website: http://www.goodbadecon.com/

DEDICATION

Writing this book was an exercise in enduring opportunity costs – many of which were inflicted upon third parties in the various countries where I visited friends only to abandon them for my keyboard. I will spend years trying to thank the following people:

At first, second and third place, Little Bell; the Very Small Australian Female who forced me to get my act together by making me food and to-do lists. Little Koala, I will never comprehend how someone who is destined to forever view the world from 152.6 centimetres can be such a towering figure. You stood your post and guarded the gates from demons on the outside and fought off those within – and made sure there were Snicker bars and cigars on my desk. You gave me time, space and a continuum – and did the pyjama dance at the same time. You cut a swathe through the thousands of little-things-to-do-popping-out-of-endless-boxes and let me have my cigar and computer time – as long as I finished my dinner. Nobody can ever love you like I do – or buy you as many watches.

To my editor; Rory, who is at the opposite end of the height spectrum. With infinite kindness, patience and common sense, you helped pull me out of a hole and never said anything unsupportive or badgering. I still don't understand how someone can know everything but it seems that you do. You also seem to find time to show me Australia in between bouts with computers and graphs. I owe you a great deal! Beer… I owe you a few beers!

To the incredible colleagues; Tom, I love you like a brother – even though my chances of getting fired increase by a factor of five every time we meet for coffee. Jon, you too are a brother – and you are such that you make Tom look good! Ryan, another brother, words cannot convey what I need to say, so I won't – but you make Jon and Tom look good! I thank all three of you CENSORED for coming by my room every so often to have fun at my expense when I am under threat of getting fired for 'inappropriateness' – or to get help stapling your ripped shirts together and to provide you with band aids (which you call 'plasters' for some reason) after your wrestling matches between classes. Katie, you are an eye in the maelstrom and we all pivot around you – especially when you serenely make Bloody Marys by the home pool! Adam, a real mate, you made sure things worked and more importantly that we worked – and you skated the fine line of diplomacy in the fracases! IB-Ian, you have been a good mate, a teacher colleague…a rock – a BIG rock! Brett, you cannot imagine what a sanctum sanctorum your humour and civility has provided – and yes, I will continue to steal your

whiteboard pens! Buncey, you are a pillar of down-to-earth common sense and cheer – and somebody I've always wanted to ask about the 'Mile High Club'! Pat, we'll crash your pub much more frequently now – get bigger bouncers.

To Ilung and Suwarni; thank you so much for taking care of me!

To my wonderful students: I dare say we have broken most of the rules at least once – but since you are grade A human beings all we did was have a laugh and get on with our work. I wrote this book with you in mind, knowing full well that few teachers in the world are lucky enough to have young people of your calibre in their classes. The smuggled-in beer for my birthday present was one of the highlights of my teaching life and that picture will be on the first page of the next edition! Under 'The Good'.

Till alla därhemma i Sverige! Ja, jag kom hem till slut…och tillbringade alltför mycket smygtid med att försöka bli klar med boken. Ni var så förstående och sade aldrig något. Vi hann ju ändå äta kräftor, fiska gös och meta bäcköring! Stort tack Guy/Ingela, Musen/Pernilla/Hermann och Lotta/Åke. Glenn och Cissi; sätt på mobiltelefonen.

To the incredible people at OSC: Kim flitted by, Keith acted dad, Jo sweated efficiency, Clara forgot her bug jacket, Kostia was…Russian, Jacek spent money on pens – and I miss Gene, Jo, George and the Oxford crowd.

Thank you Christian for my wonderful room, students, gym, food during long evenings at school and very entertaining laconic humour – and for not firing me. Thank you Julie for making sure I survived in the jungle – and for not telling Christian to fire me.

And, of course, to the Indonesian postal workers and customs officers: May you CENSORED CENSORED CENSORED CENSORED CENSORED CENSORED CENSORED CENSORED CENSORED CENSORED (by kind colleagues wanting me to stay out of prison so they don't have to cover my classes).

Matt McGee

Åkers Styckebruk, Sweden, 19th of July 2012

PREFACE

Dear Student,

It won't take you many pages of reading this book to realise that I am a very happy man. I get to do what I love, and I love what I do, which is warp young innocent minds with economics. There is nothing I want to do more than teach economics and hopefully this comes out in the very personal style of writing I use.

This book has been written for you, not your teacher. The theory content is mainstream economics and in accordance with the IB syllabus, but the style of writing is uniquely mine, and will naturally differ from other teachers in general and your teacher in particular. It is most important that you realise this, and that this book – like so many others – doesn't contain the 'truth' in any way, but merely one of many possible versions of what we call reality. You and your teachers will both agree and disagree with some of the content here, and that is exactly as it should be! From disagreement comes discussion and debate. From debate comes argumentation. And from argumentation comes learning – to both sides hopefully.

Life is fun and so is economics. In fact, I cannot separate the two any longer – much to the irritation of anyone who tries to hold a serious conversation with me. I urge you not to regard economics as a subject confined to classrooms and complex diagrams, but as an outlook on life and things happening around you. That is why I have put so many personal little stories in here; to convince you that economics is just a way of putting words to events and concepts thereby providing a little order and structure.

Two small pieces of avuncular advice: 1) Read! Read your textbook – whatever it is – and then read it again. Read newspapers and economic literature on the side. Read blogs and net reports. Read studies and essays by others. When you read economics (or about economics!) you add to your active vocabulary in the language of our Dismal Science. 2) Look things up! Never take 'truths' at face value but LOOK THEM UP. You will be truly amazed at what you find out – as will your teacher. If he/she is a good teacher then they too will look up what you have found. This is called education.

Dear Colleague,

Virtually every economics colleague I meet is either better educated, more experienced, or a better teacher than I. I have realised this for some time and therefore know that they could have written a book as good – or better – than this one. To date there are some five textbooks written specifically for the IB and you are thus not holding what is a book on economics – nor is it the book on economics. It is my book on economics, which means – for better or worse – that the examples and stories are all done from a personal vantage point, and laced with my own rather pithy sense of humour. The personal outlook is intentional, as I wish to show our students that life indeed is applied economics. The stories, jokes and shoot-from-the-hip comments are unfortunately inevitable – I simply don't find anything worth taking that seriously. Especially myself.

I sincerely hope you are not put off by the personal style of writing and come to realise that while there is lots of nonsense interspaced in the text, there is also a good deal of rather sensible and accessible economics. I have found that if one is able to hook new and sometimes intimidating concepts onto scenarios or events which people are familiar with, or can see humour in, learning becomes easier. Economics is not boring or other-worldly, so why should economics texts be that way? It is always easier to pull a piece of string across a table than push it.

Winston Churchill once said; "Criticism may not be agreeable, but it is necessary. It fulfils the same function as pain in the human body. It calls attention to an unhealthy state of things". I, my editor and proof-reader at IBID have spent countless hours seeking out errors, omissions and mistakes, yet I alone am ultimately responsible. There will naturally be mistakes in the text – either due to oversight or plain ignorance on my part. You and your students would be of invaluable assistance to me if you bring to my attention any and all errors in the text – and/or comment on the book in general. Write to me at matt@goodbadecon.com and the errors will be immediately published and commented on in the "Errata" section of the economics homepage; www.goodbadecon.com. I will answer all serious correspondence.

Yours,

Matt (matt@goodbadecon.com)

USING THIS BOOK

The structure of the book

This book follows the IB syllabus to the letter – in fact, the headings used herein are almost identical to the IB syllabus. I have taken pains to divide the syllabus into 'bite sized' chapters, 96 in all. Chapters 1 to 3 are introductory chapters with a few words on basic terms and chapters 4 to 96 follow the syllabus outline. It is definitely not necessary to follow the syllabus order (in fact, many teachers don't), yet I would recommend doing Section 1 initially, since the concepts therein are the basis for much of economic theory.

Section 1 (chapters 4 to 35) deals with the economic basics such as opportunity costs and the basic economic problems arising in societies. While this section might be considered a bit long, the intention is to give new students a breadth of examples and illustrations to make the initial meeting with economics as easy as possible. I have taken care to extensively exemplify and illustrate the use of perhaps the most important economic model; supply and demand. Having a solid understanding of how to use and apply the basic supply and demand model makes it easier to learn higher order concepts in later chapters.

Section 2 (chapters 36 to 62) deals with macroeconomics, and at this stage I have assumed students to be comfortable with basic economic concepts, so I have somewhat limited the extent of explanatory text and increased the use of case studies and applied economics.

Section 3 (chapters 63 to 77) focuses on trade issues, and is strongly linked to both micro/macro issues and development. I have put great effort into finding examples and statistics of recent date, and to use as many contextual examples as could be fit in.

Section 4 (chapters 78 to 96) is development economics and builds heavily on concepts introduced in all other syllabus sections. The 'spread' of development issues throughout the syllabus is intentional, as IB economics puts major focus on development issues and emphasizes this by relating development to a number of issues throughout the syllabus.

Throughout the book

Numerous 'Pop quizzes' have been included in every section and sub-section. These are of 'Test your understanding of the issues above' type, and to cut down on the size of the book, all the answers will be housed and periodically updated at www.goodbadecon.com.

Outside the box is where I have put non-syllabus concepts, theories and models. A little depth means that a certain syllabus concept/theory is examined further and/or applied in greater depth. Case studies bring up current and historical events pertaining to economics. Applied economics looks at specific scenarios from an economic vantage point. All of these are really just 'flesh on the bones' of the core IB syllabus content, and are included merely to tweak interest and further illustrate the linkage between our complex world and economic thought. Story time is mostly nonsense that I simply couldn't resist including. All of the content under the captions above is clearly delineated from the core body of text.

At the end of all four syllabus sections – and many sub-sections – I have included a number of short answer and extended response questions. While short answer questions are not really intended for SL, I have deliberately included short answer questions that are applicable to SL content. Data response questions have been left out completely, as I felt it better to cover a wide range of issues in as little space as possible – this is easier to do using short answer and extended response questions. Most schools have a battery of data response questions to use in practice tests and I will continuously upload a variety of these on the Goodbadecon website. Paper 3 questions for higher level are found at the end of each relevant chapter and further practice questions will be on our website.

Also on the website, there will be a section on examination criteria, command terms used in exams and exam, extended essay and internal assessment guidelines and advice, together with some samples. And, in spite of my long-standing dislike of educational practices such as 'keep it simple silly' and 'find the basics really easily', I have finally been badgered into including a glossary. Such are the forces of demand.

Contents

0

Introduction to Economics

Key concepts:

- What is economics?
- The good…
- …the bad…
- …and the economist

You know that text at the beginning of books that we all skip? Well, this is it. So, as a **Read This!** incentive, there is a secret code baked into the first three chapters with which you can login to the IBID site and win US$2,000.[1]

One of my many cheeky students was once introducing a new student around and had apparently warned the young lady about me. When they got to my class and had viewed the show for a few minutes, my student leaned over to the somewhat shocked young lady to calm her; "Don't worry. You get used to him. It's an acquired taste – it takes about three years." To which the young lady replied "But we're only in his class for *two* years!?" Evil smile from my student, "Yupp! Welcome to economics."

I'm a teacher. I teach. I love my students, my subject and my job – and all this will come across in this rather personal book for IB students. In using this book you basically become one of my students and thus get to read what all my new students in IB1 hear during the first class:

1. Economics is very cool! Yes, I know; it has a mathematical foundation…and no; very few economists have ever been arrested for tossing a TV out of hotel room after an AC-DC concert. What I mean by "cool" is that our field of study extends to virtually every aspect of society and helps to explain (going from the mainstream to the ridiculous) why taxes on imports (known as tariffs) do not protect domestic jobs in the long run and whether Bon Scott is a more efficient singer than Brian Johnson.[2]

2. Economics is about real life and real events! It is not some abstraction (= general concept) from the fuzzier parts of TOK – it is largely uninteresting if the tree makes a sound falling in the forest when nobody is there to hear it. We would instead look at how much value-added would be the case in making chop sticks out of the tree or whether the resulting soil erosion would outweigh the benefits to society of 100,000 bento box complements.

3. Putting 1. and 2. together I argue that a good grip on economic terms and concepts has numerous personal and societal benefits. Without stooping to Gekko's 'Greed is good' credo (= statement of belief) the former might deal with realising that shares in the world leading Danish windmill producer Vesta will rise due to higher oil prices and the warm fuzzy feelings that come with being an

---

1 You do realise I'm lying. Right?

2 Singers in AC-DC. You think I'm making this up? Check out "Who is the most efficient singer in AC/DC, Bon Scott or Brian Johnson?", University of Calgary archives at http://mpra.ub.uni-muenchen.de/3196/

owner of shares in renewable energy companies. As for societal issues, economics provided decision makers in the UK government with sound arguments for spending around £1.2 billion in taxpayers' money subsidising 100 Vesta windmills, which were put in place in 2010 by the Swedish firm Vattenfall in the largest offshore wind farm off the coast of Kent in the UK.[3]

4. Finally, economics is not a "dismal science"[4] but in fact a most optimistic and heartening field of study. The basic facts are that in many respects humankind is increasingly better off; we live longer and healthier lives and have immensely increased the opportunities, choices and abilities to lead better lives – noting of course that there are some notable and distressing exceptions and that the future is by no means certain.

The Good…

IB2 dinner at Hacienda Morales "Split the check amongst yourselves while I finish my cigar, ladies and gentlemen!"

The picture above is the one I use as my desktop on my computer – mostly to remind myself how incredibly fortunate I am. My IB2s and I decided that we needed a good meal so we went to Hacienda de los Morales in downtown Mexico City. Sitting around the table after dinner with a good cigar and Cognac, it struck me that this group of young people were one of Mexico's strong hopes for a better future. These "Zero-Point-Two Percenters", as I call them,[5] would basically become the movers and shakers of tomorrow's Mexico. I remembered how I introduced the topic to them some 1½ years earlier:

"Listen up people! Here's the deal. First off, economics is the study of *us;* as economists we are basically studying what we are, how we have created societies and what drives us. In studying this we are gathering information in order to outline possible solutions to very real problems and my key point here is that economics is not an abstract study (= purely theoretical or hypothetical) of "numbers and money" but an *applied science*. We look at reality with the intention of using information gathered in order to construct models and theories useful in real life, for example issues such as how to increase the production of food and goods without causing irreparable damage to the environment. Secondly, as I have said, economics is not a "dismal science"[6] but in fact rather an *optimistic* science. The basic facts are that, despite the exceptions, life has got better for most people as time has passed. My editor tell me that, when his parents married, average life expectancy was 31 years. It is now 67 years. It is the exceptions to this progress that you must deal with and thank god for a good education!"

…the Bad…

"Left hand down a bit, there's a break in the fumes ahead."

3   This is the Thanet Offshore Windfarm which came into operation in 2010. See http://www.energyboom.com/wind/thanet-wind-farm-now-online-uk-once-again-world-leaders-offshore-wind-energy, http://www.vattenfall.co.uk/en/thanet-offshore-wind-farm.htm and http://www.telegraph.co.uk/comment/columnists/christopherbooker/8025148/The-Thanet-wind-farm-will-milk-us-of-billions.html

4   The term "dismal science" was apparently coined by Thomas Carlyle in the 19th century as an argument against economist's opposition to slavery! However, the term is often attributed to Thomas Malthus of "The end is nigh" fame due to his prediction in the late 19th century that the economics of population and food guaranteed that population would ultimately exceed potential food production. No, it hasn't happened – we produce more food *per capita* today than ever before in the history of mankind.

5   As they are undoubtedly part of the 0.2% wealthiest group of humans ever to walk the face of the earth.

6   A quote commonly attributed to 19th century author Thomas Carlyle who was referring to the economists Thomas Malthus and David Ricardo.

Anybody who's read a newspaper has been bombarded by the gloom and doom scenarios which seem to sell a lot of papers. Decreasing natural resources together with increasing population gives us the core question of how the on-going depletion of natural resources and rising populations can be met by increased efficiency in the production of food and the development of alternative energy sources. The future scenarios are most uncertain but, as one single example, economists take into consideration the prospect of a production peak for oil (it already has according to the Energy Watch Group![7]) and that decreasing oil production together with the depletion of natural resources might result in *falling* food and industrial production. Other notable issues facing our world are land/soil/air/water pollution, global warming, desertification, HIV/AIDS, poverty levels, unemployment and so forth.

…and the Economist

"John Maynard Keynes, 1883 – 1946. '…a man of genius…//..who…had a world-wide influence on the thinking both of specialists and of the general public…'[8]"

So what do we do in economics? We study these issues from the point of view of how societies' wants and needs can be met now and in the future. John Maynard Keynes, perhaps the most famous of all modern economists, claimed that economics was the most difficult of sciences, as one needed to be a mathematician, philosopher, politician and psychologist.[9]

A standard textbook definition would be along the lines of "Economics is about utilising and allocating scarce economic resources to achieve optimum output and/or utility", and most of my colleagues would agree that three key concepts are at the heart of economics, namely scarcity, resource allocation and incentives.[10]

The first two concepts deal with how societies use *scarce resources* to produce the *endless needs* and wants in society. The third issue is that of how people, firms and institutions will act in accordance with these needs and wants; the *willingness* and *ability* of households to save, work and start businesses; the willingness of firms to invest, produce and innovate; and the willingness and ability of institutions to regulate such transactions and provide safety nets. In keeping with my hard-earned gunnery sergeant image, I simply say that *economics is about incentives* – everything else is a footnote.[11] I also make sure that my bright-eyed and bushy-tailed younger students get to hear me – at least once during the first week – that "…you're born, live through an endless series of trade-offs, and then die…".

Basically, economics is the study of:

1. Trade-offs or opportunity costs (money going into arms means less food)

2. Choices (why guns and not food)

3. Incentives (how do we get people to produce guns rather than food)

Chapter 2 deals with the basic economic problem which is squarely centred on the three questions above and Chapter 3 with basic data skills. These chapters are not part of the syllabus but I don't want any student of mine to be ignorant of the most basic economic issues and ways of compiling/using data.

---

7   *The Guardian*, 22 October 2007, "Steep decline in oil production brings risk of war and unrest, says new study"

8   From *The Times* obituary, 22 April 1946

9   Keynes' exact words were: '…the master economist must possess a rare combination of gifts… He must be a mathematician, historian, statesman [and] philosopher - in some degree. He must understand symbols and speak in words… He must study the present in the light of the past for the purposes of the future. No part of man's nature or his institutions must lie entirely

---

outside his regard.' (See *Essays in Biography*. *The Collected Writings of John Maynard Keynes, Vol.X*, Royal Economic Society. Published by MacMillan Press Ltd 1972.)

10   Allocation is central to economic terminology. It means roughly, 'placement', or 'the placing of'.

11   Yes, I stole this from somebody. Economics doesn't repeat itself, economists repeat each other.

A Short Story of Economic Depression...

During the Great Depression in Australia during the 1930s unemployment rate was high and many people were desperate to get work. It was common for men to walk the roads in search of work, often it was simply for their 'keep' which means food and a bed. Their bedroll was referred to as a 'matilda' or 'swag' as in the song 'Waltzing Matilda'.

The Photo shows a young man, Murray Greig, who decided to ride his bike and he also constructed a trailer to carry his belongings in search of work. He left his home in Melrose in 1935 at the age of 22 and headed east to Victoria and then north through New South Wales and then Queensland. Motor vehicles were not common and the roads were not well made and subject to flooding and other damage. He often had to carry his bike and belongings on his shoulders.

His work included tending stock, clearing land, breaking horses and tractor driving. He was also quite successful in amateur bike racing which supplemented his meagre income. He eventually arrived in Queensland in 1939 where he spent about a year cutting sugar cane before returning home. He then enlisted in the War, serving overseas and then marrying, raising a family of 5 and being a successful farmer.

... and hope!

2. Introduction to Basic Economic Terms

Key concepts:

- Factors of production
- Scarcity
- Basic economic problem
- Opportunity cost
- Production possibility frontier (PPF)
- Positive and normative statements
- Utility and marginal utility
- Micro and macro

'Sifnos'

Factors of production

We use the term 'resources' a great deal in economics, and traditionally assign all economic resources four headings; land, labour, capital and entrepreneurial spirit. These four groups constitute the **factors of production** used, to one extent or another, in the production of all goods.

Land is used in a wider sense, and covers not only the use of land for farming and space for factories, but to a wide variety of natural resources such as oil, water, timber and ore. One often uses the term raw material for these natural resources. Land is also agricultural goods such as rapeseed and fish from the sea. All of the aforementioned resources are often referred to as primary goods, (or primary commodities) and I often tell my students is that "If you can dig it up, chop it down or pluck it, it's a primary good".

Labour is pretty much self-explanatory but it is worthwhile to note that it is often the element of labour that adds value to all basic natural resources and transforms basic (simple) goods such as silicon and oil into higher value goods such as silicon chips and the plastic casing used in the computer I am using to write this.

Capital is any man-made factor of production, such as a factory or machine.[1] Yet the term is more far-reaching, as it can also

mean the whiteboard pen I use to fill the board when I am using it to produce education. Capital, as a term, is specific in that the item must be used in the production of goods and not in simple consumption. One could say that capital is defined more by usage than anything else. A guitar in my hands is simply a consumer good (and a mistake) while in the hands of Lady Gaga it is a capital offence…sorry, that's 'capital'.

The **entrepreneur** is the person who brings the other three together and creates goods to fulfil wants and needs in society. From Edison's light bulb to Picasso's painting Guernica all production necessitates the idea, drive, and ambition of an entrepreneur to put land, labour and capital together and create something. As a personal addendum, I would add education, training and experience to the four production factors above – this is commonly referred to as 'human capital'. Human capital is frequently considered of increasing importance in modern production.

> **Definition: 'Resources/factors of production'**
> The **factors of production** are commonly divided into land, labour, capital and entrepreneurial drive. These are the resources necessary to create/supply goods and services in an economy.

or resources – just try to build a boat on a deserted island with a stack of Yen!) 'Capital' is one of many terms with subject-specific meaning which also has other, wider, meanings outside of economics. In its purest usage in economics capital is a man-made factor of production, yet in a wider more general way we use it to stand for financial or physical assets which can generate income, such as property or shares in a company.

---

1   Be very careful in using the term 'capital' in economics! Very often the term is confused with 'money' – which is NOT a factor of production. (Money is just a representation of goods

Scarcity

In outlining the factors of production (land, labour, capital and the entrepreneur) available to the economic system, we inevitably come upon the issue of **scarcity** which has a rather specific meaning in economics as it goes to the very heart of what economics is all about, namely the optimal use of resources. Scarcity means that all societies face the common problem of limited resources and how to best allocate these resources to provide for our endless wants.

Scarcity is a universal problem, but does NOT necessarily mean that all peoples in all economies lack the same things! A most valuable resource is water, which the Swedes have in abundance in Sweden but has to be shipped in by tanker every day to the Greek island of Hydra where I lived as a child. On the other hand, on Hydra there is no lack of master stone-masons and marble, enabling marble sinks, counter-tops, tiles – we actually had a toilet seat made out of marble![2] When I think of it now, the amount of marble we had there could have funded my university studies.

This is my point; scarcity is an issue for all nations and has been for all time. It's simply a matter of what is scarce and the reasons for this relative scarcity. Scarcity is defined by availability of resources, true, but also by our wants and desires, which are infinite. No matter what need is fulfilled, there is always another lurking in the background and this is true in all people. Naturally all these needs cannot be satisfied as there are limits to society's ability to satisfy them, the reason being that while our needs are seemingly *endless* the resources (land, labour, capital…) used in satisfying them are quite definitely *finite*, or limited. Scarcity is what one might call a triumph of harsh reality of the inborn wants of man; all societies during all ages will have wrestled with the abundance of human wants, the inability of the economic system to supply all wants, and the resulting choices resulting from scarcity of available resources.

> **Definition: 'Scarcity'**
> The excess of wants resulting from having limited resources (land, labour, capital, and entrepreneurs) – scarcity arises in satisfying the endless wants of people using limited resources. **Scarcity** is a universal problem for all economies – it is not limited to "poor" countries.

Basic economic problem

The **basic economic problem** that economics seeks to address has permeated all societies throughout history:

- *What* to produce?
- *How* to produce it?
- For *whom* to produce it?

The first issue, '**what**', deals with the allocation of resources to make the goods that society wants. The issue can be as trivial as 'red shoes or blue shoes' or as broad as the classic 'Guns or butter' question we shall look at in just a moment. The issue of '**how**' deals with production methodology, organisation and technology. The final issue, '**who**', is the wider issue of distribution, i.e. to whom the spoils of production go. All societies have to deal with these issues, from the Aztec society which created the lookout tower up the hill here on Expat Street in Mexico City, to the centrally planned economy of Cuba which made the excellent cigars I – but not the diners at the table next to me – enjoy.

> **Definition: 'The basic economic problem'**
> The enduring central issue of economics; how all human societies, throughout time, are forced to deal with the questions of **what** to produce, **how** to produce it and for **whom** to produce.

Opportunity cost

Resources are scarce and societies' wants are endless; this means making a choice which in turn means giving something up. An **opportunity cost** arises as soon as one alternative means giving up the *next best* available alternative. For example if my preference ranking (in descending order) in spending $US10,000 is 1) a Blancpain watch; 2) 4 weeks vacation in Cancun; 3) a new car, then in choosing the Blancpain my opportunity cost is 4 weeks vacation – not the new car, because my opportunity cost is the highest ranking – e.g. *second best* – alternative I give up.

---

2   Eh, a bit nippy during the winter.

Definition: 'Opportunity cost'

Opportunity cost is the option foregone in making a choice of "Alternative A" over "Alternative B". Assuming that all possible choices have been ranked in order of preference, **the opportunity cost is the relinquishing of the second best possible alternative**, i.e. the next best foregone opportunity of obtaining the highest ranked of all possible alternatives. The concept is fundamental in all subsequent economic concepts.

The issue of 'what/how/whom' is naturally one of choices. In some way, one has to choose between options. Let us use the learning tool of choice for Aristotle; the *syllogism*. A syllogism is a set of factual statements ordered in a state of natural progression which all lead to an (inescapable?) logical conclusion. The classical syllogism is one where Aristotle referred to his teacher Socrates:

1. 'Man is mortal'

2. 'Socrates is a man'

3. 'Socrates will die'

The economic syllogism is somewhat less dramatic:

1. People's (or society's) wants are infinite

2. Resources are finite

3. Choices must be made

We'd best comment on the above, primarily in order to explain and perhaps defend the premises therein.[3] Most people would ultimately agree that, no matter what one has attained, one *always* has an additional want. Mick Jagger couldn't get no satisfaction (and knew You Can't Always Get What You want) – and he was a drop-out from the London School of Economics! Fulfilling one's desire for a new watch doesn't mean that one wouldn't want a vacation in Cancun and a new car to drive there in. Even Bill Gates, who has earned millions in interest alone during the time it has taken me to write this section, will have unfulfilled wants.[4]

The premise of endless wants and thus hard choices holds as true for society as for the individual. A municipality (= local

3   We will be using the term 'premise' rather often. A premise (pl. premises) is a basic assumption – often in a line of reasoning or argumentation where the premise must be included at the outset in order for subsequent conclusions to hold true. For example, my premise in writing this is that my students are adequately versed in English.

4   Detractors may say that 'world domination' comes the closest.

government) might face the decision of allocating funds towards new computers for the local public school or an all-expenses-paid fact-finding mission to Monaco for municipal councilmen and spouses. National government might have a choice between 15 new fighter jets and a new cancer research centre. To phrase it in economic terminology, there are simply too few resources (e.g. land, labour and capital) available to enable all wants to be taken care of. We are thus presented with interminable *trade-offs* at personal, municipal and national level: the Blancpain or vacation; school computers or vaca… em…fact-finding mission; fighter jets or health care. A trade-off means mean an opportunity foregone.

Production possibility frontier (PPF)

All economies face the same issues outlined in the basic economic problem of "what to produce…how…and for whom", and all economies face infinite wants and limited resources.

Assume a small island economy, Sifnos, and the following:

1. the economy only has the resources to produce *two goods*, pottery and tourist services

2. there is a *known maximum output* and the economy can attain this level – i.e. the economy can fully utilise all resources so there will be no unemployment or idle machines

3. there is *no trade* with other villages, thus production equals consumption (much more on this in Section 4)

4. that any given quantity of resources transferred from one sector to another are *unequally productive*, (or re-allocatable) – all resources used in the production of pottery are simply not equally productive in the production of tourist services.

Figure 2.1 shows a few of the possible combinations of output in the Syfnos economy. The PPF is drawn assuming that all factors of production are fully utilised whereby any combination of output *within* the PPF (**point E**) shows that the economy is producing at a sub-optimal level, e.g. that there is unemployment, idle capital or unused natural resources. **Points A** to **D** make up the boundary of possible output (maximum efficiency – see *Pareto optimum*, Chapter 7), and **point F** is outside the PPF and thus impossible to attain.

Figure 2.1 PPF for Sifnos – increasing opportunity costs

Figure 2.2 illustrates the issue of opportunity cost. Output ranges from 600,000 hours of tourism and zero pottery to 5,000 tonnes of pottery and zero hours of tourism. Let us say that Sifnos is at **point A**, where the Sifnians are producing 600,000 hours of tourism and nothing else. Moving to **point B** to produce the initial 1,000 tonnes of pottery entails giving up 20,000 hours of tourism service, which is thus the opportunity cost of the first tonne of pottery. The striking difference for the Sifnians will occur when they increase output of pottery further, from **point**

B to **C** along the PPF. It is immediately obvious that the cost ratio has changed – the second 1,000 tonnes of pottery entails giving up the production of 40,000 hours of tourism service. In other words, the opportunity cost of the second 'batch' of pottery is 40,000 tourism hours – twice the cost of the first 'batch'. Moving from **point C to D** we see how the opportunity cost of pottery production rises to 140,000÷1.8 ≈ 78,000 hours of tourism per tonne of pottery.

Figure 2.2 PPF for Sifnos – increasing opportunity costs

Opportunity costs occur in economies because resources are scarce. *Increasing* opportunity costs in production are due to the fact that all resources are not equally re-allocatable or, to use another phrase, **factors are not perfectly mobile**. When the Syfnians re-allocate limited (scarce) resources from tourist services to pottery the best available resources (factors) in the production of pottery will be used first. As additional factors are taken from tourism, the economy is forced to use increasingly less productive resources in the making of pottery and thus will have to take increasing amounts of resources out of tourism production. Each additional unit of pottery will result in increased opportunity costs.

In Figure 2.3 (I) at right, the economy shows **economic growth** since actual output has expanded from **point A** to **B**.[5] Our model economy has just increased the production of both capital and consumer goods while the PPF remains in place. This example of growth illustrates that unused resources have been put into use For example, firms start utilising unused machines and hire previously unemployed labourers. Note that we are assuming that the quantity and/or quality of factors of production remain the same so the PPF has not shifted. **Points A to B** could illustrate an economy recovering from recession – previously unemployed labourers are hired and idle machinery is put to use.

Remember, the PPF is a purely hypothetical construct; we have already taken into account all possible production circumstances such as technology, labour skills and quality of raw materials in drawing the curve. That is why more or better *use* of existing available resources illustrates growth in the PPF without an associated shift in the PPF. Figure 0.3I. and Figure 0.3 II illustrate the effects on both potential output (the PPF shifts outwards) and *actual output/growth* (**point A'** to **B'**). The increase in productive potential and subsequent actual growth is the result of more and/or improved resources being put to use.

Diagram (I)

> Increased use of unused factors moving the economy towards the PPF – the economy shows economic growth.

Diagram (II)

> An increase in the quality or quantity of available factors of production increases potential output (PPF_0 to PPF_1), and as these factors are put into use the economy grows (point A' to B').

Figure. 2.3 Economic growth – two possible illustrations

Changing the *quality* of the factors of production

An increase in the *quality of labour*, such as education, training, and experience, will inevitably result in the ability of labourers to produce more goods during any given period of time.[6] The *quality of land* can entail new farming methods, high-yielding crop types, or purer iron ore for making steel; improved *capital*, such as production technology could render more widgets[7] per hour and

---

5   Note that 'point to point' movements in PPF diagrams are grossly exaggerated. In Figure 2.3 I above, it appears as if both consumer and capital goods have increased by 25%. Again, keep in mind that the diagram is only a description and not a scaled representation of reality.

6   Yet anyone having spent a summer working on an assembly line, as I have, knows that one uses one's experience and skill to have longer breaks – not to produce more! I gleefully note that the firm in question went bankrupt.

7   A **'widget'** is a fictitious/hypothetical product or 'thing'. Economists use the term 'widget' to make up for a sad lack of fantasy and imagination in trying to conjure up good real-life

possibly also use less raw material to produce the same amount of widgets; and innovation in the production of capital goods in turn means improved machines for consumer goods.

Changing the *quantity* of the factors of production

If there are *more* factors of production available, then potential output has increased. Using available *capital* in the manufacture of more capital increases potential output; adding more *land* (as the Dutch have done for hundreds of years by way of building dikes) adds to potential arable land; and an increase in the *labour* force (for example a "baby boom" generation exiting school or increased immigration) all mean that potential output has increased – and that *actual output* increases when these resources are employed.

Positive and normative statements

Positive statements have nothing to do 'positive attitude' or such. A *positive statement* is one which does not contain subjective or opinionated elements but is based on facts that can be proved or disproved. Examples of positive statements are 'Socrates is a man' (see my 'economic syllogism' in the section on opportunity costs); 'He was sentenced to death'; 'Socrates was a convicted paedophile'; or perhaps even 'Socrates was a woman'. Note that nothing value-laden is included. It's just stating facts – or ridiculous assertions as the last two are intended to show – that can be tested by looking at data and facts as far as they are known. Therefore 'positive' in no way means 'good', but rather that there is no value judgement involved. So if I were to say 'There are very few repeat-offenders amongst those who have been executed' I am actually using a positive statement.

However, stating that 'The death penalty is abhorrent (= objectionable) to a modern democratic country' is a *normative statement*, just as if one – like those who convicted him – were to say 'Socrates was immoral'. The reason is that one would be standing on a platform of *evaluation*. Often this type of statement is referred to as being 'value-laden', i.e. being burdened with moral/ethical values. One is using a certain standpoint – religious, cultural or philosophical – to express the value of something that cannot be refuted by logic or objective reasoning. One is using a *normative* vantage point to express opinion.[8]

> **Definition: 'Positive' and 'normative'**
> **Positive statements** are based on facts or evidence, free from value-laden and subjective standpoints. They can be proved or disproved using a scientific approach. **Normative statements** are based on norms, thus they are subjective and biased – they can not be proved or disproved. "Man evolved from apes…", a Darwinist proposition, is a positive statement while "Man was created by god…" is clearly a normative statement.

The aim of economic theory is to provide a basis for analyses which is as unburdened by normative statements as possible. Generally speaking, scientific research should be confined to positive questions – i.e. questions or hypotheses which can be either *verified* or *falsified* by looking at real life observations and data. The stricter we are in our use of definitions, assumptions and consequent models, the more we can weed out the more emotional aspects of economic conclusions. Economics as a science attempts to outline 'good' and 'bad' without resorting to subjective views. An economist aims to pose questions that can be stated positively by defining 'good' and 'bad' in economic terms rather than moral terms and then answering the questions by examining facts and evidence.

Utility and marginal utility

Utility refers to the *usefulness perceived* and *satisfaction derived* from the consumption/use of a good. The perception of a good's utility forms our preferences. The term springs from the Latin 'utilitas', meaning usefulness or applicability. Just think of 'utensils'![9]

> **Definition: 'Utility'**
> **Utility** is the benefit/satisfaction/usefulness one gets from the consumption of a good. The concept is quite obviously highly normative as peoples' preferences vary greatly. My perceived usefulness will not be the same as yours.

Utility is one of the classic *abstract concepts*, right up there with 'quality' and 'happiness'. (In fact, I once bet one of my students $US100 that he could not define 'quality' and he never got

---

products in textbook examples.
8   A norm is a rule or guideline arising from within the evolved standards of a society, e.g. "Socrates should be punished by being forced to drink poison!"

9   This translation was provided years ago by Pia Birgander, my incredible IB coordinator in Sweden. She knows about 18 languages, Latin being but one. Now, having her around is utility – talk about useful!

back to me. He has now graduated from the London School of Economics!). It simply defies any form of enumerative measurement and appraisal, just like 'love' – try grading your love for your boyfriend/girlfriend on a curve! The same goes for utility; one simply cannot put the satisfaction one might feel, say, from the consumption of chocolate into numerical values. Picture yourself standing in front of long row of shop windows filled to the brim with goods. You are standing in front of a certain shop window – and other people are standing in front of other windows. Why isn't everybody at the same shop front?!

Simply, utility differs between prospective customers and this steers preferences and therefore selective perception. While perhaps you are fogging up the window of the Italian shoe shop, your friend might be smearing nose grease on the armoured glass in front of the Swiss watch shop.[10]

---

10 Take it from an old person: don't shop in pairs! The opportunity cost of looking at things which bore you out of your skull quickly becomes unbearable when you realise that you are giving up your own valuable shopping time.

Story-time!   "You did what?!"

This was voiced loudly and almost in concert by a group of colleagues who had gathered for evening grape juice (= wine) and yummies at one of the houses along Expat Street here in Mexico close to the school we all work in. Word had apparently spread that I'd bought a new watch for some ungodly sum and Marc – being a stout Yorkshire lad with no inhibitions whatsoever – handed me my wine, glanced at my watch and said "Ay, looks good! I used to see those all over the place in the black markets in Vietnam".

"You most certainly did not", I growled. "This isn't a back-alley copy but the real thing – proudly manufactured in Geneva by the Little Swiss Watch Gnomes."

By this time my friends in the know were all grinning like possums eating the core out of a Mexican cactus, waiting for Marc's look when he found out how much it cost. I can't quote him in print, but basically he said "You're crazy!" It took about 3.4 seconds for this to become the consensus view and if their looks had taken on physical action I would have left the party in a straitjacket. Virtually everyone had different views on "…the insane waste of money…" and "…what one could have gotten instead…". The alternatives ranged from a new car to 4 weeks at a good hotel in Cancun or Acapulco. What they also all agreed on was that I shouldn't wear it anywhere "risky" – which in Mexico means "…anywhere outside the house and/or near the police".

The point of this story is that the satisfaction I feel from wearing (e.g. consuming) the watch is different from what the others would feel. I derive such pleasure in owning fine mechanical timepieces that I am prepared to pay a great deal for them – and thus give up numerous other goods. In economic language, I am spending my income in such a way that I get maximum pleasure from the goods I spend my money on – I am maximising my utility. Any other option would have added less to my overall sense of satisfaction. By allocating my income towards the consumption of a (-nother) watch, I added to my "happiness" the most. This is marginal utility, dealt with next.

Utility is at the heart of economic reasoning when it comes to analysing peoples' wants and purchasing habits. As we shall see, utility explains why we buy a good in the first place – but the concept does an even better job of explaining why we buy goods in the *second* place! Say what? 'Second' place? Well, stop and think; have you in fact bought many single goods in your life-time? Are we not, in fact, constantly buying more goods, i.e. another good rather than a good?! If so, then it should be of far greater interest for the social scientist to explain why we buy an additional unit of a good rather than the first one, seeing as how recurrent purchasing is far more common than a 'virgin' purchase. Point in fact; I wasn't buying *a* watch (see Story Time below) but *another* watch. Thus we need a concept which addresses the *addition* to utility caused by consuming one more unit of the good. This is **marginal utility**.

> **Definition: 'Marginal utility'**
> The addition to total utility (i.e. total benefit/satisfaction/usefulness or well-being) resulting from the consumption of one more unit of a good. Ultimately choices are based on the perceived marginal utility of one more unit of a good.

My favourite example of marginal utility – and one which lends itself to immediate recognition and knowing laughter by every student I've had – is the 'cold beer in the desert' allegory (= comparison). Picture yourself marooned in the Kalahari with nothing but the clothes on your back and your trusty credit card securely tucked away in a Velcro pocket. You set off on foot towards what you hope will be the nearest town. It's 45° C in the shade – which would be fine if there was any shade.[11] But you are walking in the mid-day sun, which is hot enough to fry the niblets off a bronze elephant.[12] Walk, walk, walk. Sweat, sweat, sweat.[13]

Suddenly, through the shimmering heat waves and consequent mirages, you see a large parasol in the distance. You do a high-speed shuffle towards it and coming closer discover that it is real – you also see the painted sign with the legend 'Mabogunje's Finest Kind Cold Refreshments! All major credit cards accepted!' There are a number of bar stools set out in front of a bar counter. You set yourself down on one of the stools in front of the smiling Mabogunje, slap your credit card on the bar and order a beer. Mabogunje pulls the draft-lever a few times and sets an ice-cold pint of beer in front of you.

You start slow and end fast – just like a good opera. Indeed, quaffing away there, you actually hear blissful choirs and feel like the injured Tristan being cured by Isolde. What sweet satisfaction! 'It ain't over 'til the Fat Lady sings', you think and order another beer…

Whap! Mabogunje slaps another cold one on the counter and whips your card through the register. Your initial thirst having been satiated (=satisfied), you drink the next one with a bit more ease. It too is incredibly tasty – though it doesn't quite give you the lust-filled experience of the initial stein. Still, it's satisfying enough for you to want a third…which is, again, very tasty but not quite in the same satisfaction-league as the previous. You feel better and better after each additional beer – but not at the same rate. This continues through the 4th, 5th and 6th beer. Each additional beer will add to your overall satisfaction – and inebriation – **up to the 7th beer**.[14] Putting this little story into the traditional illustrative method of the economist, we get tables and diagrams.

Quantity	Total utility	Marginal utility
0	0	
1	100	→100
2	170	→ 70
3	215	→ 45
4	242	→ 27
5	257	→ 15
6	262	→ 5
7	259	→ –3

Figure 2.4 Table of total and marginal utility

The table in Figure 0.4 above shows how total utility increases – but that each additional beer adds to total utility less than the previous beer. Shakespeare was well aware of marginal utility when he wrote "Can one desire too much of a good thing?" The

---

11 The correct term for '°C' is degrees Celsius – not 'centigrade'. The scale was invented by the Swedish scientist Anders Celsius in 1742. (Source: http://www.astro.uu.se)
12 I dare not go into the subject of niblets. You are much too young and innocent.
13 Or 'Perspire, perspire, perspire' if you are being polite.

14 This is when you inadvertently start crawling in the sand violently sick or fall into a drunken stupor. Or both – in that order.

first beer renders total utility of 100, the second beer renders total utility of 170 – thus there is *diminishing marginal utility* since the second beer adds to total utility by 70 rather than an additional 100 units of utility. Note that one shows 'movement' – i.e. rate of change – by putting any marginal values *between* the absolute values. For example, the second beer increases total utility from 100 to 170 which is shown in the table and in Figure 1.8 below as a marginal utility value of 70 between the total utility for beers number one and beer number 2. Figure 2.5 below shows how total utility increases – at a decreasing rate, i.e. diminishing *marginal* utility – up to the 7th beer. Swigging the 7th beer confers *negative utility*, i.e. *disutility*, and you wind up grovelling in the sand, proposing marriage to a cactus or picking fights with claw-equipped nocturnal animals. Basically you'd be prepared to *pay* someone to drink the 7th beer!

Marginal utility has wide applicability in the study of demand patterns. It helps to explain consumption habits and patterns and also why goods such as diamonds – which are in no way necessities – are very dear, while water – which is an essential of life – is nowhere near as expensive as diamonds. I shall return to this classic economic conundrum in Section 2.2.

Micro and macro

Microeconomics centres on the forces working at the individual level, e.g. the individual firms' and consumers' (often bunched together in households) behavioural patterns and decision making processes. Here economics focuses on the needs, desires and buying habits of the individual consumer in conjunction with the output capabilities of firms for particular products. In short, microeconomics looks at firms' outputs and pricing decisions and consumers' purchasing decisions, for example studying how firms react to increasing costs of production by raising the price and subsequently how consumer/household spending is adjusted when the price rises.

Withdrawing from the study of individual market participants in order to study the broad interaction of the aggregate (= sum, combined, cumulative) of separate (micro) parts is the purview of **macroeconomics**. The four main issues here are aggregate output (and thus economic growth), price level (and thus inflation), labour markets (e.g. unemployment), and finally foreign sector dealings (such as the balance of trade and exchange rates). Central to the study of macroeconomics is the business cycle (also known as the trade cycle), which shows total output in the economy over time – this is often put into the context of variations in economic activity (recessions and expansions) and the links to macroeconomic policy.

The diagram graphs the total utility in consumption of beer. Each additional beer adds to total utility – up to and including the sixth beer.

The diagram graphs the marginal utility of consuming an additional beer. The first beer renders 100 units of utility, while each additional unit (pint) adds to total utility less than the previous – this is diminishing marginal utility. The seventh beer actually decreases total utility, i.e. it renders negative marginal utility of –3.

Figure 2.5; Total and marginal utility of drinking beer in the desert

Definition: 'Micro economics'

Microeconomics is the field of economics dealing with the relationships between individual components in the economy; firms, industries and consumers (households). This interplay is the basis for individual markets.

Definition: 'Macro economics'

The sum total of all micro parts, the aggregate of individuals' and firms' behaviour. The four main areas of study are; 1) growth (increase in total output); 2) price level (inflation); 3) labour market (unemployment) and; 4) the balance in the foreign sector (exports/imports, exchange rates).

Let us compare the two using a few examples:

Looking at how *a firm* reacts to increased demand for its product is a *micro-issue* while studying the effects on *all firms* in the economy due to a general increase in demand is a *macro-issue*; the decision of *a worker* to work less due to lower wages is *micro* while *total hours* of labour (and unemployment) is *macro*; the effects on *an industry* (= group of firms producing similar goods) due to higher labour taxes is *micro* while the effect on *total production* in the economy due to taxes is *macro*; government legislation aimed at *monopolies* is *micro* but government legislation aimed at increasing taxes on profits for *all firms* is *macro*.

The distinction between the two areas is admittedly blurred, somewhat contrived and ultimately none-too-useful. I often find the two areas overlapping and would be hard put to draw a concrete line between them in many real life situations. Basically the difference isn't vitally important unless an exam question addresses the issue… which is just the sort of statement designed to get me into a hefty argument with some of my colleagues.[15]

Summary & revision

1. Economics deals with societal wants and the ability to provide these wants using scarce resources (factors of production – land, labour, capital and the entrepreneur).

2. Economics deals with the trade-offs we face every day, ranging from the question of going to the movies or doing homework to the use of government funds for producing weapons or schools.

3. Resources are scarce since it is impossible to satisfy endless societal wants.

4. The basic economic problem is what to produce, how to produce it and to whom the goods will go.

5. For every use of resources there is an alternative use – this gives rise to opportunity costs, i.e. the second best alternative foregone (quantity of Good B) in the production of a given quantity of Good A.

6. The PPF shows the possible combinations of output in an economy. The boundary (frontier) is where all factors are fully utilised. A shift outward of the PPF (economic growth) is caused by an increase in the quality and/or quantity of factors of production.

7. Positive statements are statements which can be proven or disproven. There is no evaluation or value-laden element in such a statement.

8. Normative statements are subjective. They cannot be proven or disproven.

9. Utility is the benefit or satisfaction derived from consuming a good. Marginal utility is the satisfaction derived from the consumption of one more good.

10. Microeconomics looks at the individual level of firms and households while macro looks at the aggregate (sum) of all households" and firms' activities such as unemployment, inflation and economic growth.

---

15 I answer all correspondence. Please write to me at matt@ibid.com.au

3. Introduction to Basic Data Skills

Key concepts:
- Models
- Correlation
- Causality
- Nonsense (spurious) correlation
- Indexes
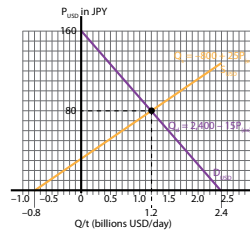
You will have to deal with data in economics, both in writing your economic analyses for internal assessment and in your exams. Internal assessment comprises 20% of your total grade and the data response questions 40% at SL and 30% at HL. This section takes a look at how models are built, some common types of economic data and how these are frequently presented in tables and diagrams. One main objective here is to get you to understand the difference between correlation and causality as these concepts are very useful in evaluating/criticising data which has been compiled (= put together) for you in tables and diagrams. Another objective is to clearly outline the difference between data based on stock and flow values. Finally, since so much economic data is given in index form to enable comparisons between countries and over time, basic index series will be looked at.

Models

Models in economics are based on observations of real life and then trends seen in data collected.

1. The first step is to organise reality in a rational manner by **observation** and subsequent gathering of data.[1] It is most important to define the concepts that are being systematized; if we are studying income and the effects on consumption then it is imperative to clearly outline whether we are looking at gross or net income, including social benefits or not, what time period, which population…etc.

2. Step two is spotlighting aspects of the data in order to see whether any **patterns** emerge. Any and all models and/ or theories which emerge will inevitably be limited in applicability, but the basic strategy is to find *correlation* (= a relationship or correspondence) between the variables being studied. We might see that consumption increases as net incomes rise, e.g. positive correlation between the two variables.

3. Finally, one formulates a **hypothesis** based on any patterns. For example, this might be; 'Increased income results in an increase in consumption'. One then tests this by applying the model to other sets of data (such as the income and consumption figures over a longer time period or for another country) in order to see whether the model has more general applicability. This means testing whether the results are generally consistent with the theory being formulated. A key issue here is if variable X causes a change in variable Y…or vice versa. This is the issue of *causality*.

Correlation

When a set of data shows that there are visible and perhaps predictable patterns in the variables in our data, we speak of correlation. Note that high(-er) values for income and corresponding higher(-er) values for consumption is *positive* correlation, while high(-er) values for income and low(-er) values for birth rates would show *negative* correlation.

For example, one of the models used in macroeconomics is the **Phillips curve**, which shows how inflation (rise in the general price level) and unemployment (percentage of working population not holding a job) are related over a period of time. Up to the 1960s in the US there was strong *negative* correlation; when unemployment (U) rose inflation (i) fell…or maybe I should say that when inflation rose unemployment fell!? In

---

1 It should be noted that many a cynical social scientist has observed that 'data is the plural of anecdote…if the anecdotes have been selected systematically'.

economic shorthand we would pose this as '$\Delta U \rightarrow \Delta i$' and '$\Delta i \rightarrow \Delta U$' respectively.[2]

Causality

This is a common conundrum (mystery, problem) in economics and indeed in all the social sciences; what causes what? Here, does the change in unemployment *cause* the change in inflation or the other way around? This chicken or egg problem is known as *causality*, which basically looks at the relationship between variables and seeks to highlight which is the **cause** and which is the **effect**.[3] Let us look at a simple example.

Nonsense (spurious) correlation

It is important to note that correlation is NOT the same as causality! Consider three variables; income, consumption and birth rate. Standard economic theory shows that an increase in general incomes in a country will lead to higher consumption rates as any given percentage of income going towards consumption will necessarily mean higher consumption as incomes grow. We have *positive correlation* between the variables income and consumption. Studies also show that as incomes grow (on a national level) the birth rate markedly falls – this means that there is *negative correlation* between income and birth rates. Hence, a collection of data on birth rates and consumption might show how birth rates fall when consumption in households rises – this would in all likelihood be an example of 'spurious' or *nonsense correlation*.

Thus, while there is correlation in all three cases above, one would be on thin ice indeed in claiming that there is also *causality* in all pairs. In the case of income and consumption most economists would agree that the increase in income causes an increase in general consumption – a relatively clear example of cause and effect. Most studies would also show that increasing incomes over time leads to a drop in birth rates (often measured as 'live births per 1,000 women') as higher tax revenues provide social welfare systems and pension funds which in turn enable couples not to have to rely on having a good many children to take care of them in times of need and old age. Rising incomes will cause lower birth rates. In other words, there is also a causal

flow (causality) in regards to changes in income and subsequent changes in consumption and birth rates.

The point to be made now is that if we were to study the variables 'consumption' and 'birth rate' we might well find that there would be negative correlation; as total consumption rises over a period of years, birth rates fall! This is where a good, sceptical outlook on the use of statistics is absolutely vital. It is pretty obvious that increased consumption does not cause lower birth rates or vice versa, but that the two are in causal contact with a common underlying variable; income. Figure 3.1 below illustrates the difference between the variables that are correlated and the variables which show causal links.
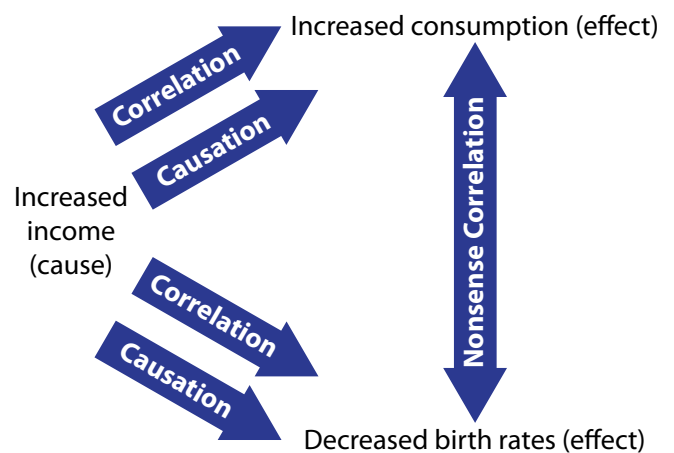
Figure 3.1 Nonsense correlation

Correlation without causality is 'spurious' (= 'fake', 'false') or 'nonsense' correlation. There are some pretty far out examples of studies which – in varying degrees of self-seriousness – find strong correlation between a number of variables. One classic study in 1875 by the famous economist William Jevons found correlation between sun-spots and the business cycle (see: http://cepa.newschool.edu/het/profiles/jevons) while more recent (jesting?) studies show positive correlation between the amount of churches and violent crime! (See: http://www.selu.edu/Academics/Education) My personal favourite, however, is the classic study from 1950 showing an astoundingly strong (positive) correlation between the number of people classified as 'mental defectives' and the number of radios in households during the 1930s and 1940s.[4] (See: http://www.sommestad.com/text/essay)

In these three examples, the answer to seeming correlation can be found in underlying, common variables. Sun spots might cause changes in weather patterns and thus crops which

---

2   In fact, the answer is probably neither! It is, instead, the underlying change in overall demand for goods and services that causes both inflation and unemployment to change. See Chapter 54.

3   While this might seem rather trivial and straightforward, I assure you it is not! Some of the most heated debates and schisms (divisions, splits) between economic schools of thought have arisen over this issue.

4   They must have been listening to an early version of rap music. Proto-rap, maybe.

ultimately could lead to increased harvests and falling grain prices and, in turn, trigger a crisis. Churches are more frequent in highly populated areas and this often means cities – which have higher crime rates than the countryside. As for radios and correlation to mental illness, the study was carried out during a period of rapid societal change which would lead to the radio becoming a common household item and vast improvements in diagnosing mental illness. Thus, in all three cases, seeming correlation must in fact be explained by entirely different data.

Final note on correlation/causality: As any number of my colleagues delight in pointing out, it is a mainstay of the failings of the dismal science of economics that it is often impossible to actually definitely *prove* causality.[5] Too often there are so many underlying and parallel variables involved that it becomes impossible to attribute (or even show) causality between a pair of variables. There are simply too many other possible influences to attribute a degree of causality to any one variable. I strongly recommend Chapter 4 "Where have all the criminals gone?" in the book "*Freakonomics*"[6], where the authors claim that the remarkable decrease in crime rates in the US during the 1990s had little or nothing to do with a "roaring economy", harsher gun laws, increased use of the death penalty or new police strategies but was clearly linked to the famous US abortion law (*Roe v. Wade*) from 1973! The authors make a case that mothers who are socially disadvantaged – single parents with low income – are strongly correlated with having children turning to crime. The dramatic increase in abortions during the 1970s meant fewer births in the very social groups where children were at worst risk of becoming criminals. The effect of this was that by the time the early 1990s rolled around, there were far fewer early teens with the propensity to become criminals and the crime rate began to fall. I hasten to point out that the authors very quickly question their own results by stating the "…likeliest first objection…", namely that there might NOT be causality but simply correlation. They *then* use several statistical methods to test their findings and indeed find causality.

Definition: 'Correlation'
When there is a visible pattern between variables (X, Y, Z…), one speaks of correlation. If $\Delta\uparrow X$ and $\Delta\uparrow Y$ there is positive correlation. If $\Delta\uparrow X$ and $\Delta\downarrow Y$ there is negative correlation.

Definition: 'Causality'
When there is correlation and evidence that a change in one variable causes a change in another variable, there is causality, e.g. if ΔX causes ΔY.

Definition: 'Nonsense correlation'
If there is correlation without causality, one speaks of nonsense or spurious correlation.

DEALING WITH EXAM QUESTIONS
In data response questions you are often given tables, charts and diagrams. One neat way to show evaluation is to comment on any indications of correlation and causality. For example, in a set of time-series diagrams for an economy showing how export earnings (X) have increased, inflation (i) has increased and national income (Y) (money value of all goods produced in an economy) has increased, one can use core economic theory to comment on the correlation and even causality: "…as export revenue is part of national income, increased exports cause an increase in national income and the increased demand for exports together with a causal link between income and consumption will drive up domestic prices in the economy, i.e. inflation rises…"

Indexes

An index is a very clever way of adjusting data so that it becomes easy to compare changes over time and/or between individual firms/industries/countries etc. Two of the main indexes you will deal with in economics are for inflation (the consumer price index, CPI) and comparisons of economic growth rates (increase in GDP) between countries. It's worth knowing how an index is put together at the outset of studying economics, so here is a simple – and childish! – example.

---

5   One of my favourite colleagues, L.M., has the occasional HL math lesson in my classroom. Once, I was having a rant about the Mexican traffic jams and the bungling traffic cops who showed up and actually made matters worse. L.M. said; "Matt, you're confusing the causal flows! It's not the traffic jam that *causes* the cops to show up – it's the traffic cops showing up who *cause* the traffic jams!" I don't win a lot of points with L.M.

6   *Freakonomics*, <u>Levit</u>, Steven D., and <u>Dubner</u>, Stephen J., 2005

In my childhood home in a place called Åkers Styckebruk,[7] Sweden, there is a well-scribbled and well-hacked doorpost where countless children, grandchildren, friends, and neighbourhood kids have marked their height during the past 30 years. Careful scrutiny yields the name of my oldest friend, Guy, several times and quite frequently Amanda – his daughter. *Tempus fugit*[8] and all that. Here's what Amanda's marks on the doorpost look like:

Figure 3.2a Growth rate and index

We see four specific points in time in Figure 3.2 above and Amanda's height at each specific time. The change in height between the four points in time is, of course, growth. The rate of change per time period is illustrated in a simple table below.

Time	Jan. 1996	Jan. 1997	Jan. 1998	Jan. 1999
Amanda's *height*	111 cm	114 cm	123 cm	133 cm
Change during period (i.e. growth)	Period 1 '96-'97 3cms			
		Period 2 '97-'98 9cms		
			Period 3 '98-'99 10cms	

Figure 3.2b Amanda's growth 1996 to 1999 in absolute values

If we want to compare Amanda's growth rate to other kids over the same period of time, it's easiest to index her growth. Using the first time period as the base year – the value that all coming values will be compared with – we calculate:

$$\frac{\text{Height at } t_n}{\text{Height at } t_0} \times 100$$

…where is the actual year and t_0 is the base year value. This gives us Figure 3.3 below.

Time	Jan. 1996	Jan. 1997	Jan. 1998	Jan. 1999
Amanda's height – indexed (Jan 1996 as base year)	100	102.7 $\left(\frac{114}{111}\times100\right)$	110.8 $\left(\frac{123}{111}\times100\right)$	119.8 $\left(\frac{133}{111}\times100\right)$

Figure 3.3 Amanda's growth 1996 to 1999 – indexed

Note that the rate of growth over the entire time period is 19.8 percent. However the percentage change *between individual years* is not a matter of deducting the index value of the previous year from the value of the year in question. For example, the percentage increase between 1998 and 1999 is not 9% (119.8 – 110.8)! Percentage changes is calculated by taking the change in index value and dividing by the original value. So the *percentage increase between 1998 and 1999* is [119.8 – 110.8] ÷ 110.8 × 100. This gives us a 8.1% increase in height between 1998 and 1999.

Using Sweden's GDP figures as a final example for the same time period we can calculate growth rates using both absolute values and then an index series[9]:

---

7   Not pronounced 'Acres of Stickybricks'.
8   'Time flies'. (And, according to my mother, gravity sucks.) Amanda is now in her 20s and I cringe every time I get a message from her on Facebook. Uncles really don't want to know what those cute little girls turn into when they grow up.

9   Swedish Bureau of Statistics, at http://www.scb.se/templates/ tableOrChart____26651.asp

Applying the same methodology as to Amanda's growth.

Year	1996	1997	1998	1999
Real GDP for Sweden (SEK = Swedish crowns)	1,958 Billion SEK	2,002 Billion SEK	2,070 Billion SEK	2,143 Billion SEK
Change during period (i.e. growth)	Period 1 '96-'97 2.24%			
		Period 2 '97-'98 3.40%		
			Period 3 '98-'99 3.53%	

Figure 3.4 Economic growth rates for Sweden.

Example: period 2; $\frac{2{,}070 - 2{,}002}{2{,}002} \times 100 = +3.40\%$.

Naturally the index series shows the same growth rates as when we use the absolute values. Again, the cunning thing about using indexes is that we can more easily compare Sweden's growth rate over time and between, say, Sweden and the UK. I always urge my budding economists who write extended essays in economics to index data values in order to be able to compare more easily. One of my more recent essays dealt with the possibility of correlation between economic growth and quantity of marriages in a municipality. By indexing all values it became much easier to do an analysis of the strength of correlation.[10]

Summary & revision

1. Economists use lots of models. These are based upon observations of real life; data collection; correlation analysis; formulating hypotheses; and finally testing the model against reality.

2. **Correlation** means that two (or more) variables show a pattern. For example we know that there is negative correlation between the price of a good and the quantity sold over a period of time.

3. **Causality** means that there is correlation and also that a change in the value of one variable causes a change in the value of the other. If the price of ice cream decreases this will cause an increase in ice cream consumption.

4. When there is correlation but no causal relationship, one speaks of nonsense or "spurious" correlation. There is 100% correlation between eating food and dying (ultimately) but eating food does not cause death.

5. A very common method in economics to enable comparisons over time and between groups is to index all data. The base period or value commonly has the value 100 and all subsequent values of the index are in comparison with the base value.

---

10 This is called regression analysis and is best done having a math teacher at your elbow or on speed dial. Incidentally, there is indeed correlation betwen economic growth and quantity of marriages! Positive or negative?!

1.1

4. Markets and Demand

Key concepts:

- Definition and function of markets
- Law of demand and correlation between price and quantity demanded
- How individual demand comprises market demand
- Introduction to basic market model; downward sloping demand curve
- Movement along vs. shift of demand curve
- Non-price determinants of demand

HL extensions:

- Linear demand function; $Qd = a - bP$
- Shifting the demand curve, e.g. a change in "a"
- A change in the slope of the demand curve, e.g. a change in "b"

Syllabus section 1 (Chapters 4 to 35) deals with how markets attempt to solve the basic economic problem. Consumers' wants are paired-off with firms' ability to satisfy those wants, creating the keystone of economic analysis: the supply and demand model. Then we look at market failings such as over-consumption or under-consumption of goods such as tobacco and health care respectively. Higher level students will then go into greater depth concerning market outcomes and the decision-making process within firms in 'Theory of the Firm'.

Definition and functions of markets

Pick up the morning paper and flip to the classified ads. 'Used bicycle – call Adam.' 'Baby-sitter urgently needed! Call the Svenssons!' 'Kittens! Call Sandy.' 'Change money at Sami's – no commission!' All of these fictitious ads are examples of markets in action. In each and every case, there is something being offered or asked for. Any replies will thus show the flip side of the coin; someone who wants the good or someone who can supply the good.

Markets are well-established institutions and operate at all levels of human interaction. When one thinks of 'markets', one often envisions stalls in an open square where goods are put out for sale and potential customers meander looking and perhaps buying. This traditional market place is a brilliant example, as all the necessary prerequisites are in place; there are numerous buyers (or customers) and numerous sellers (suppliers). Many of the stalls will have similar goods and similar prices yet many people will still circulate in the market, perhaps foregoing the wares offered at one stall in order to look at the same goods at another. Is that haggling I hear?! '..one dinar for this?! How about two dinars for three?!'

Definition: 'Markets'

A **market** is a situation where potential buyers are in contact with potential sellers. It enables the needs and wants of both parties to be fulfilled whilst establishing a price and allowing an exchange to take place.

Market place at Chichen Itza, Mexico. Maya-Toltec style, 900-1200 AD.

Cut to an air-conditioned office equipped with wall to wall computer screens and phones. 30 young men and women are glued to the screens – all the while talking in to multiple phones and feverishly taking notes. Cyber-dating? Nope. Just another day in the life of («**blip**» and AUD$250,000 were just transferred to Tokyo) a foreign currency exchange office.

You get the picture. An example could be from yesterday or from the Vikings celebrating their successful invasion of York in 866; "Yah, Sven, plenty women to kill and cattle to rape! I'll trade you one ox for two women."[1] A more recent example is from any modern financial institution dealing with international buying and selling of currencies on the Forex (foreign exchange) market. The only real difference between the two markets' structures is that the buyers and sellers of currencies are not in physical contact with each other – and also that the Australian dollar is not bought at the business end of a 4.5 kilo broadsword.

Market types

In the description of markets above it is implied that there are a number of firms competing on the market for goods and services. While this is often enough the case, there are a number of instances where the degree of competition is considerably lower, for a number of reasons. The *competitive* market has many firms and a high degree of competition, whereas a *monopoly* is at the other extreme; a single powerful firm and no competition. In between there is a 'large fuzzy grey area' where we find markets characterised by a few large firms – *oligopoly* – and high profile firms characterised by heavy advertising and brand-imaging – *monopolistic competition*. The following figure arranges these four market structures according to four sets of criteria. (For greater depth on market structures – even for the ambitious SL student! – see chapters 22 – 35.)

---

1   Or something to that effect. They weren't too brainy those Vikings. Too much salted herring.

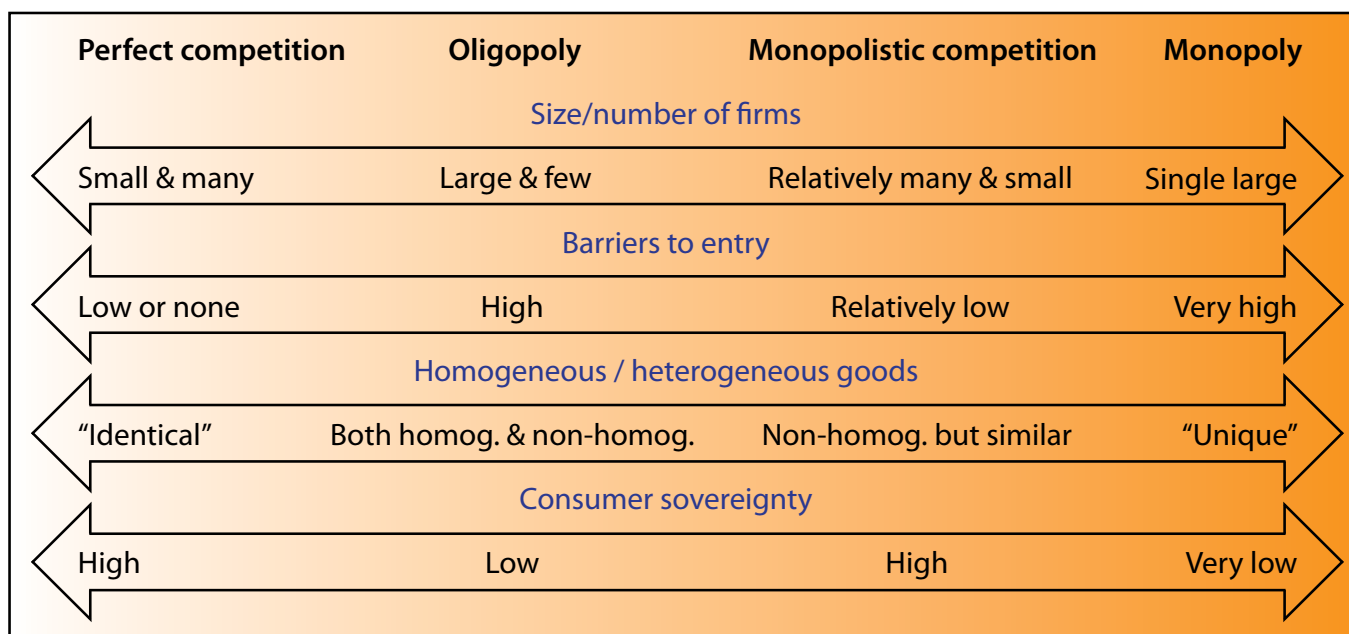| Perfect competition	Oligopoly	Monopolistic competition	Monopoly
Size/number of firms			
Small & many	Large & few	Relatively many & small	Single large
Barriers to entry			
Low or none	High	Relatively low	Very high
Homogeneous / heterogeneous goods			
"Identical"	Both homog. & non-homog.	Non-homog. but similar	"Unique"
Consumer sovereignty			
High	Low	High	Very low

Figure 4.1 Market structures and their characteristics

The four criteria in Figure 4.1 above are in fact used as assumptions, as we are defining the market structure by assuming that certain traits are fulfilled. The **size and number of firms** has a powerful influence on the level of competition and price, and therefore the power of the firm relative to the consumer. Many firms will mean more competition and vice versa. The greater the number of firms that compete, the less power any individual firm will have over market prices.

Barriers to entry mean that it is difficult/costly for potential newcomers (firms) to enter the market. High barriers to entry are mainly of three types:

1. **legalistic** (for example the restrictions and regulations governing banks and airlines, or not having certain patent rights necessary for production);

2. **financial** (including difficulties in getting the funding together to start a railroad company, or not having access to raw materials); and finally

3. **economic** (newcomers would be producing small amounts which will mean high cost per unit whereas existing firms would have much lower costs). The various forms of entry barriers compound the question of competition, as high barriers to entry will allow existing firms to act without giving too much consideration to the possibility of firms entering the market and increasing competition.

The degree of **homogeneity** and **heterogeneity** defines the composition or make-up of the good. Homogeneous means 'same' or 'identical' and in economics signifies whether a good has any number of identical (or close) substitutes or not; the potato is an oft-cited example. Heterogeneous is of course the opposite; a good which is differentiated (= set apart, non-homogeneous) from possible substitutes. It is relevant to note that goods do not actually have to be physically different – it is sufficient for us to perceive that goods are different – there are many Taco restaurants but only one *Mama Rosita's*! (See monopolistic competition below.)

Finally, there is a question of the relative bargaining power of consumers and firms. If consumers have a good deal of influence on prices and output, then **consumer sovereignty** is high. One would indeed expect to find this in competitive markets where there are many firms. If, however, there is only one firm (supplier) then there are no substitutes available and consumer sovereignty will be low. Regarding the empowerment of consumers as the ability to 'cast votes' on goods, then a competitive market will empower consumers more than non-competitive markets.

The shading in Figure 4.1 is not by accident. If you look at the characteristics of the perfectly competitive market and monopoly, you will see that they are each others' opposites, being basically two extremes. Oligopolies and monopolistically competitive firms, on the other hand, portray very little correspondence in terms of neatly following the arrows separating the two extremes. This 'fuzzy grey area' is the most difficult to characterise accurately and unfortunately this is the area most of the real world winds up in. Also, don't be fooled by the neat figure; in reality there is a great deal of overlapping, where an oligopoly is monopolistically competitive and could also have elements of monopoly markets embedded. An example would be Microsoft, Macintosh and Linux operating systems.

Perfect competition

A perfectly competitive market is characterised by many buyers and many sellers, all interacting in such a way so as to provide the highest possible quantity at the lowest possible price. As the goods are assumed to be 100% homogeneous, the only competitive element is price-competition, which in turn empowers consumers and the market is demand-driven. In providing this outcome there is no waste – all goods are produced in order to fulfil market demand. This optimum outcome in terms of resource allocation is what is 'perfect' about the market structure. **Agricultural goods** like tomatoes and coffee and **basic commodities** such as iron and copper could be considered perfectly competitive market goods. The remaining three market structures are commonly referred to as 'imperfect competition'.

Oligopoly

The main defining elements of an oligopoly are 'few' and 'large' – where there could well be several hundred firms but four or five firms dominate the market. The dominant firms in an oligopolistic market structure might have access to limited raw materials, such as bauxite (for aluminium) and oil, creating **entry barriers** for other potential firms. Firms are often large because of the necessity to produce very large quantities of goods to cover high costs of production and research and developments (R & D) – such firms are said to enjoy **benefits of scale** (scale means size) where the cost per unit of output falls as the firm increases in size. Examples of such oligopolies are pharmaceutical and car companies. There is also an incentive for firms to **collude** (= cooperate), for example by agreeing to

set identical prices or by dividing the market up geographically so as to avoid head-on competition. This type of behaviour is frowned upon in most industrialised countries and severe legal penalties are often set for collusive behaviour.

Monopolistic competition

This is an increasingly common market structure where there are a large number of firms producing similar goods which are **differentiated**. The defining elements of monopolistic competition are taken from perfect competition and monopoly – hence the name. The market is **competitive** as entry barriers are low and potential firms will have access to attractive markets. The market is also **monopolistic**, as goods are highly profiled – firms put a great deal of resources into marketing in order to convince us that while there are many possible substitutes, there is only one 'Brand X'. This is known as **branding** and serves to create in the mind of the beholder that a particular good or service is in some way different - e.g. superior - to others. Standard textbook examples of monopolistically competitive firms are restaurants, hotels, car repair services and bakeries. One also often finds a number of markets which are largely monopolistically competitive but where there are a few large, dominant firms; the markets for sports shoes and soft drinks are notable examples. In reality therefore, the line between monopolistic competition and oligopoly is often blurred.

Monopoly

Finally, the far end of the spectrum. Here the assumption in economic theory is that there is one firm only – a pure monopoly. Examples of *pure* monopoly markets are postal services, state television and even oil companies (Pemex in Mexico) as seen in many countries. The concept of a pure *private* monopoly doesn't stand up to scrutiny in the real world particularly well and I'd be hard put to find a specific example.[2] The definition of monopoly is a tricky one since it often depends on the definition of the market in terms of both *product* and *geographical area*. Hence, de Beers might be considered a monopoly (see footnote below) in the supply of raw diamonds but not in the market for diamond jewellery. Geographically speaking, local gas, water and cable companies can have a regional monopoly but not

2   Diamonds are a *possible* example, where DeBeers has for over 70 years controlled some 50% to 90% of the total market for raw diamonds – and the European Union, threatening a fine of up to 10% of de Beers global sales, prohibited the company's method of controlling the market via intervention purchasing in 2006. (*International Herald Tribune*, Feb 22, 2006; "De Beers loosens grip on diamond market"). Much of the definitional difficulty here arises due to the complexity of the market. HL will return to this issue in theory of the firm.

on *all* gas, water and TV/internet services in the country. The definition becomes a question of the *degree* of market power for the firm, i.e. that monopoly firms have market power at the expense of consumers as there are no close substitutes and thus scant competition. Firms have the power to set output and price without taking competitive forces into consideration. In the final analysis, consumer sovereignty in a pure monopoly situation is a bit like voting in the ex-Soviet Union; everyone could vote freely but there was only one party.

Law of demand

Part of the economic problem is that people have endless wants. However, simply wanting or desiring a good does not constitute demand. Demand is more an activity than a state of mind, e.g. when you are actually willing to purchase a good at a certain price. Demand in market terms is the quantity consumers are **willing** and **able** to buy at a given price, not what they would like to have. I would most assuredly want a Patek Phillipe wristwatch but I am not in the market for one (yet if you bought this book I'm getting there). When consumers are both able and willing to buy a good, economists speak of *effective* demand.

Recall from Chapter 3 the issue of correlation and causality. If we plot out effective demand for a normal good with price on the *Y*-axis and quantity demanded on the *X*-axis, we would see **negative correlation:** when the price of a good falls the quantity purchased of the good will increase. The starting point for our model of demand is positing (= proposing) that a rise in the price of a good will lead to lower quantity demanded for the good – keeping all other variables constant. This is the **law of demand**, which states that *'Ceteris paribus'*, **a change in price will lead to a change in the quantity demanded'.**

> **Definition: 'Law of demand'**
>
> *Ceteris paribus*, a fall in the price for good X will result in an increase in the quantity demanded for good X. An increase in the price of X will result in a decrease of quantity demanded.
>
> The law of demand states that as long as all other variables (income, price of other goods, preferences) remain unchanged, then a fall in the price of a good will lead to an increase in the quantity demanded. This means that the demand curve is downward sloping and thus that quantity demanded is negatively correlated with price.

Ceteris paribus

In scientific research in general and economic models in particular, one builds a model knowing that there are many possible influences on the relationships one is looking at. In using our demand model to look at the effects of a change in the price of, say, gasoline (petrol), we would assume that a number of other influences on gasoline consumption would be kept constant. Consumption of gasoline would be correlated to several non-price determinants, such as households' *income*, the average price of *cars* (complement goods), the average price of *public transport* and so on. Assuming that there is no change in income, car prices or public transport, we would see that a fall in the price of gasoline would lead to an increase in gasoline consumption.

The assumption of **ceteris paribus** – 'all else equal' – is almost always present in our economic models. It is virtually impossible to leave out of our models as it raises the level of scientific trustworthiness by creating a more rigid framework of deductive reasoning, e.g. that a change in 'X' will lead to a predictable change in 'Y' as long as the situation is not muddied by an infinite number of other possible influences.

> **Definition: 'Ceteris Paribus'**
>
> *Ceteris paribus* is Latin and means "…all else remaining equal…" or "…all other things remaining the same…". This essential assumption allows economic models to predict outcomes and relationships with a degree of certainty and conviction simply by *assuming that variables not addressed in the model are kept constant*.

Individual demand and market demand

One of my favourite movies[3] as of late is *Despicable Me* which at the time of writing has just left the big screen and hit DVD/Blu-ray. I missed it on the big screen so,[4] like a goodly many of my students, I bought it on DVD. My students and I all have our *individual demand* for DVDs and the sum of all demand here in Mexico City makes up the *market demand*. In other words, any market demand is simply the aggregate of individual consumers' demand for the good. It is noteworthy that individual buyers can be Matt and his students (consumers), Wal-Mart (other firms) and the Naucali Youth Centre (government).

If we hugely simplify the market and look at a small market consisting of three IB students and their demand for DVDs over a week, we get Figure 4.2 below. (MXN = Mexican pesos)

The market demand curve in this very simple example is made up of the three consumers which all have an effective demand for DVDs; Maria, Ana and Jesus. In keeping with the law of demand, we see that a lower price will mean greater quantity of DVDs purchased each week. At a price of MXN150, the three consumers will purchase a total of 8 DVDs per week – which is market demand (D_{mkt}). At a price of MXN50 market demand is 24 DVDs per week.[5]

3　Pause here for groans from my students…"…Matt, *all* your movies are favourites!" Yes, so?

4　I saw it on a long distance flight from Hong Kong at 03:00 in the morning. I am sure the other passengers found my screams of laughter very amusing.

5　Yes, I know this is hideously simple. When I explain this the first time in class, there's always one of my kids who goes "Duh!" However, keep in mind that the Maria's, Ana's and Jesus' weekly allowance doesn't change…nor does the price of movie tickets…nor does the price of DVD players…nor anything else which would affect weekly DVD demand of my students.
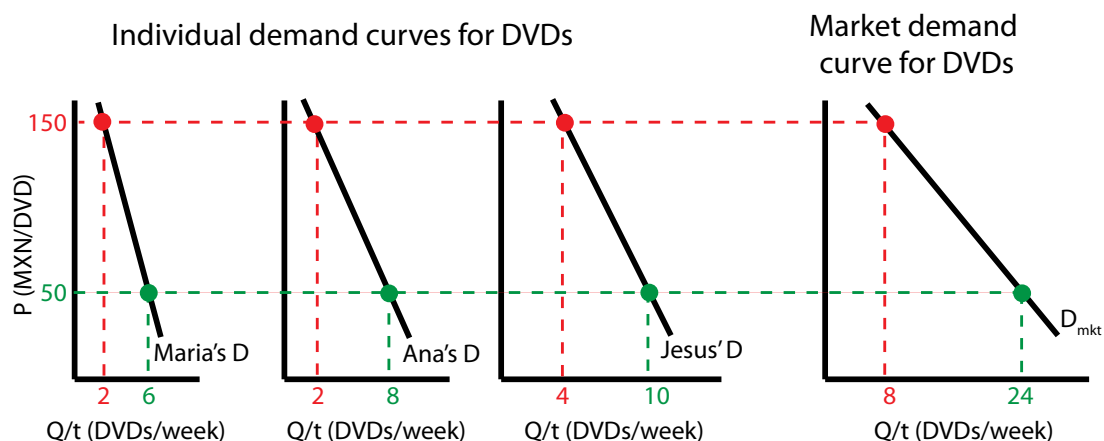
Figure 4.2: Individual demand and market demand for DVDs in Naucalpan

Downward sloping demand curve

In essence, we are painting a picture of the *pattern* of demand for a good, where all possible prices are coupled to the quantities demanded. This gives us any number of possible combinations of price and quantity demanded of DVDs, but for the sake of simplicity in Figure 4.3 below only three prices are mapped out. There are two reasons for an increase in quantity demanded when the price of a good falls; the **income effect** and the **substitution effect**.

- **Income effect:** If your income remains the same but prices rise, your real income has decreased and consumers would have the propensity (= tendency) to decreases their purchases of the good. In the same way, a fall in the price of a good means that people who have a demand for the good are now richer in *real terms*, i.e. their income in terms of how much they can actually buy has increased – they buy more. If – *ceteris paribus* – the price of DVDs falls from MXN100 to MXN50 the quantity demanded increases from 16,000 DVDs per week to 24,000.

- **Substitution effect:** Secondly, a fall in the price of a good means that *relatively* speaking the price of alternative goods has increased, so people will change their purchasing behaviour and switch some expenditure to the (relatively) less pricey good. In simple terms, consumers will *substitute* other goods with the lower priced good. Thus, if the price of DVDs increases from MXN100 to MXN150 then consumers will, to a certain extent, switch to substitutes (say cinema tickets or Blu-Ray discs) since these substitutes are now relatively cheaper – quantity demanded falls from 16,000 DVDs per week to 8,000.

> **Definition: 'Income and substitution effect'**
>
> **Income effect:** If, *ceteris paribus*, the price of a good falls, then consumers' real income (income in terms of how much they can actually buy) has increased, so consumers buy more of the good.
>
> **Substitution effect:** If, *ceteris paribus*, the price of a good falls, then the price of the good in relation to other goods (i.e. substitutes) will cause an increase in consumption. (Relatively speaking the price of other goods has increased.)

> An increase in the price from MXN100 to MXN150 causes a movement along the demand curve from A to B. This is termed a decrease in quantity demanded.

> A decrease in the price from MXN100 to MXN50 causes a movement along the demand curve from A to C. This is termed an increase in quantity demanded.
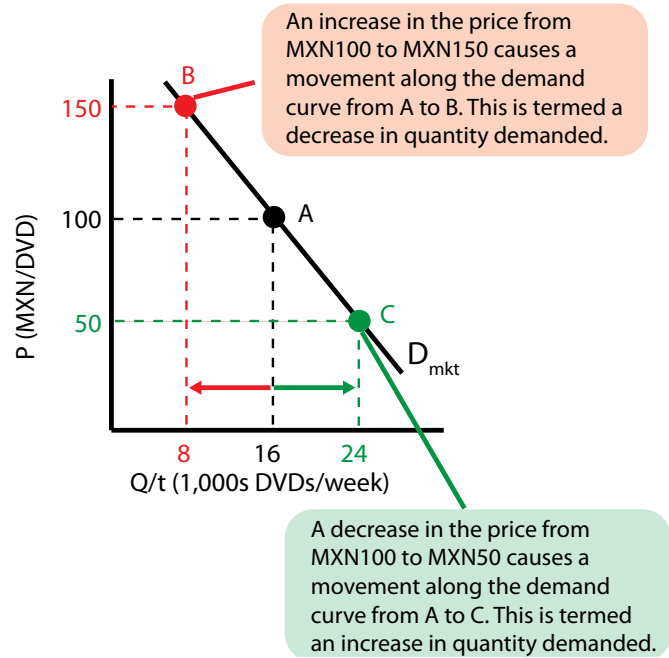
Figure 4.3: Market and downward sloping demand curve

Movement along vs. shift of demand curve

In Figure 4.3 above, assume that the initial price is MXN100 and thus that quantity demanded per week is 16,000 DVDs. If the price increases to MXN150, then the quantity demanded falls from 16,000 DVDs per week to 8,000. This is shown by a movement along the demand curve, from point A to B, which means, using correct economic terminology, that '…the quantity demanded decreases…'. A decrease in the price from MXN100 to MXN50 is a movement along the demand curve from point A to C – the quantity demanded has increased from 16,000 DVDs per week to 24,000. The demand curve extends or contracts, i.e. any movement along the demand curve can be referred to as an extension/contraction of demand. The curve is actually created by all the combined points of price and quantity demanded.

A few comments on the simple relationship in Figure 4.3:

1. The **relationship between price and quantity demanded is negative**; any downward sloping curve shows that an increase in one variable (price) leads to a decrease in the other variable (quantity) which is entirely in keeping with the law of demand. Note once again that in "moving along" the demand curve we are assuming that all other influences on demand are held constant, e.g. the *ceteris paribus* condition.

2. Have you noticed that we have put the independent variable on the **'wrong'** axis?![6] The norm is to put the independent variable on the *X*-axis and the dependent variable on the *Y*-axis. As quantity is undoubtedly the dependent variable, we have 'switched places' in our model.

3. It is most important to keep firmly in mind that there is a **time-frame** involved, here '1,000s of DVDs per week'. Economists are notoriously sloppy and lazy for some reason when it comes to including this rather important premise, and I urge you to include the time element in your own diagrams. Not having a time frame (i.e. a time limit) would basically mean that the demand curve is set in stone for all eternity! It is not; the demand curve is, as we shall see, most dynamic over time.

Shifts in the demand curve

As will be explained in some depth below, any change of the determinants of demand other than price, will serve to shift the demand curve.

A decrease in the price of Blu-ray discs (a substitute, i.e. a non-price variable of demand) leads to a decrease in demand for DVDs. This is shown by a leftward shift of the demand curve.

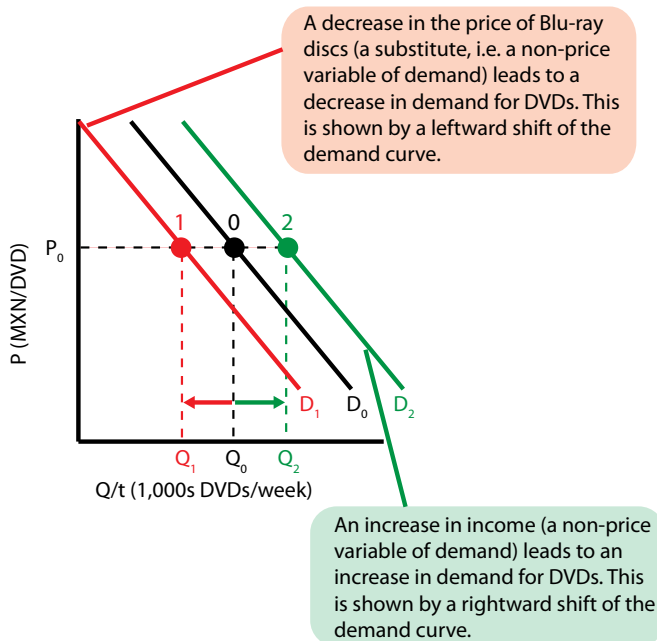An increase in income (a non-price variable of demand) leads to an increase in demand for DVDs. This is shown by a rightward shift of the demand curve.

Figure 4.4: Market and downward sloping demand curve for DVDs

In Figure 4.4 above the original demand curve is D_0. If incomes increase (for those who are part of effective demand) we would see that demand for DVDs increases at *all* price levels – this

is shown as a rightward shift in the demand curve from D_0 to D_1 and referred to as an *increase in demand*. Conversely, if the price of Blu-ray discs were to decrease, then some consumers would switch some expenditure to Blu-ray discs (a substitute) and the demand for DVDs would decrease at *all* price levels – the demand curve shifts left from D_0 to D_1. This is a *decrease in demand*.

Exam tip; using the language of an economist

Ever wondered about the language that language teachers insist upon using? Me neither. Rather than saying 'do-words' they say 'verbs'; rather than 'thing-words' they say 'nouns'. They have an entire menagerie (= collection) of terms that could just as soon be said in Martian as far as I'm concerned. Unfortunately, economists are no different – something I will never admit to the face of a language teacher. We have an immense array of very subject-specific terms and concepts. The successful economics student must incorporate them into his/her active vocabulary. Distinguishing carefully between 'change in the quantity demanded' and 'change in demand' is an excellent place to start.

Any change in the price will result in an increase or decrease in the quantity demanded. An increase or decrease in price means a movement along the demand curve. This is referred to as a change in quantity demanded.

- changing the price does not change demand but the quantity demanded
- a change in the quantity demanded is a movement along the demand curve

Any change in a non-price variable will cause an increase or decrease in demand at all price levels. An increase or decrease in demand means a shift of the demand curve. This is referred to as a change in demand.

- changing a non-price variable (income, price of other goods…etc) does not change quantity demanded but demand
- a change in demand is a shift of the demand curve

---

6 This is the fault of one of the most influential textbook authors in economics, Alfred Marshall (1842 – 1924). The story has it that he simply made an error – which has stuck with us ever since!

Non-price determinants of demand

In the coming model, a fundamental premise is that demand is a **function** of many different variables, the main determinant being the price of the good. All other influences are **non-price determinants** of demand which will *shift* the demand curve, the main ones being;

1.   **income** of consumers

2.   **price of other goods** (substitutes and complements – more on this later)

3.   changes in consumers' **tastes**/preferences

4.   changes in consumers' **expectations**/ hopes

5.   **population** changes – both in terms of size and structure

6.   **derived demand** – e.g. where the demand for cars creates demand for steel

> ### Definition: 'The demand function'
> "$Q_{Dx} = \int Px; Py; Y;....n$ " The demand function reads "The demand for good X is a function of 1) the **price** of good X; 2) the price of good Y (substitute or complement); 3) **income** (Y); 4) **n** – being tastes, population…etc. Our diagrammatic analysis links the *price* to quantity.

This book doesn't permit going into the other 1,800 possible influences on demand. Let us look at the six non-price determinants above using a few real-world examples. Note that in using the demand model, actual quantities and prices are seldom used. The diagrams serve as illustrations rather than real-life depictions of actual data.[7] (Figure 4.4 below covers all six examples used – note that the axes would change depending on the good! Swiss watches would have Swiss Francs per watch on the *P*-axis and thousands of watches per unit of time on the *Q*-axis, whereas Caterpillar machinery would have average price per machine on the *P*-axis and number of machines per time unit on the *Q*-axis.)

A decrease in the price of Blu-ray discs (a substitute, i.e. a non-price variable of demand) leads to a decrease in demand for DVDs. This is shown by a leftward shift of the demand curve.

An increase in income (a non-price variable of demand) leads to an increase in demand for DVDs. This is shown by a rightward shift of the demand curve.

Figure 4.5: Market and downward sloping demand curve for DVDs

Income

It cannot come as a surprise that an increase in income would increase demand for a good. Just recall that our wants are endless! Thus, any furthering of the ability to satisfy those wants will mean that people will have an increased tendency to do. When demand increases for a good due to an increase in income, economists commonly refer to such goods as 'normal' goods. (See income elasticity in Chapter 11.)

I do two weeks of teaching at the Oxford Study Courses in England every Easter.[8] The first time I did this I flew via Stanstead Airport, which is a small and somewhat basic airport unlike the larger Heathrow or Gatwick Airports. On returning home I told my then wife[9] that we were lucky that I flew via Stanstead, as the wages I'd been paid were burning a hole in my pocket; had I flown via Heathrow I probably would have spent the entire week's wages at the Tax Free shop on a new watch. Sad but true. So, did I then save my increase in income? Nope. I was surfing the Internet for Blancpain watches that very evening.

---

7   Though it is *strongly* recommended that you include in your diagrams any numerical values given in the articles you will use in internal assessment.

8   Basically, I'm greedy – and Blancpain watches aren't given away.

9   I seem to collect not only wrist watches but ex-wives. One of my cheekier students plotted out my marriage habits and then correlated my ups and downs with the business cycle. She found correlation – the regression coefficient was 0.35! Maybe you should write me and ask about my marital status before buying any shares or property… Hmm, good extended essay material here. Write me and I'll send you the data.

What might cause an increase in general incomes in an economy? Perhaps a decrease in income taxes leaves people with higher disposable incomes. Or general increasing prosperity gives the people in an economy higher wages and thus incomes. In any event, an increase in disposable income, (= income after taxes and including any welfare benefits and such), will increase the demand for normal goods. In the example of Swiss watches, this will cause a shift in the demand curve to the right, from D_0 to D_2 in Figure 4.5, i.e. more watches are demanded at all prices. (SFR = Swiss francs.)

I'd best add a brief note on 'abnormal' behaviour within the context of the income effect. The goods exemplified above are **normal goods** as the norm (= custom) is for demand to increase when income rises. However, it is possible that demand for some goods might actually fall when incomes rise simply because a rise in income will change households' preferences – increased income could cause households to substitute certain goods with other, more preferred goods. Such goods are called **inferior goods**, possible examples of which are public transport and potatoes. We return to this issue in Chapter 11.

Price of other goods

Substitute goods: When two goods are in competitive demand they are said to be substitutes – and thus a rise in the price of one good would cause an increase in demand for a substitute. There are a number of new formats for recording and viewing movies at home, the most recent are the formats Blu-Ray and HD DVD which look to compete with the standard DVD system.[10] When

one looks at these products, they are obviously competitors for my money. Should the price of Blu-Ray players increase there will be an increase in demand for Blu-Ray players (D_0 to D_2 in Figure 4.5) Note that, once again, we are assuming all else equal, i.e. there is no change in the quality or function of either good. Consumers are simply substituting one good with another.

Other examples of substitute goods are rail travel and bus travel, apples and pears, or, to be most product-specific, Pepsi and Coke[11] . (Many textbooks give tea and coffee as examples, which I find to be utter nonsense – but then I am heavily addicted to coffee. Nothing substitutes coffee in my book.) A recent example of the power of substitutes arose in 2010 when demand for US grain rose more than 14% due to drought and subsequent price increases in two of the largest grain producing countries, Russia and Ukraine.[12]

Complement goods: Now, how would one go about buying one of the DVD or Blu-Ray players? Being an economist, you would look at the total ("bundle") price of usage, in other words the price of the player *and* the discs – either for recording or playing. These two goods are in **joint demand**, commonly referred to as **complement goods**. The two are a 'package deal' basically, as one is useless without the other. Being complementary to each other, a change in price of one good will affect demand for the other. Should Blu-Ray discs increase substantially in price, one can expect demand for Blu-Ray players to decrease (D_0 to D_1 in Figure 4.5).

Other standard examples of complement goods are tennis balls and tennis racquets, film and cameras, and staples and staplers.[13] Note that all complement goods are often not as 'joined-at-the-hip' as my example, but are as often as not quite weakly connected via preferences and habits.[14] I often give the examples of Cognac and cigars, strawberries and cream, and mustard and pea soup. I bet that last one caught your attention. Yes, in Sweden

---

10 I actually wonder if Blu-Ray will take off – the picture is too good! A goodly number of my students agree that the picture sharpness is a bit staggering. I think my initial comment was something like "It's too real…like viewing the world sober. I'm not sure I need more reality."

11 My students disagree with me most adamantly here! Me, I failed the 'Pepsi Challenge'.

12 "Wheat, Corn Stockpiles Dwindle After Russia's Drought" BW 25 – 31 Oct, 2010

13 How about tequila and writing economics?! Try reading the footnotes….

14 Here in Mexico, a bill clip holding one's driving license and a 100 peso note (circa USD9) are complement goods – if you need to show your license to one of Mexico's finest you'll also need a "*mordida*", i.e. a bribe to get you on your way. Commonly the license and the bill are handed over together. Also, drinking and driving are not substitutes but complements. This is the sort of comment that apparently caused my colleagues to start betting on my life expectancy here. Nobody has me down for more than four years. When the colleagues say goodbye to me on Fridays, they say it with depth and mean it.

one customarily puts a dollop of mustard in one's bowl of pea soup. I added this example to illustrate the cultural dimension of many complement goods – just think of French fries with ketchup (USA), French fries with vinegar (England) and French fries with mayonnaise (Belgium). Here in Mexico, lime juice is a complement to…well, anything in the way of food![15] Oh, so is chilli![16]

Tastes and preferences

> "The Atkins diet was a nightmare for the consumer, but also for us." (Guido Barilla, chairman of Barilla Pasta, *Economist*, Jan 20th 2007)

Consumers are often persuaded to buy certain goods by peers, fashion and advertising – and I'd hate to say which of these is the strongest in terms of the Atkins diet. What if we instead look at the amazing life expectancy of the Greeks who have one of the highest average life spans in Europe – in spite of having one of the highest European proportions of smokers in the population and one of the lowest levels of physical exercise! The answer, according to research, lies in the use of healthy olive oil. A successful advertising campaign would increase demand for olive oil, shifting the demand curve to the right (D_0 to D_2 in Figure 4.5). Our tastes can also change as the result of intellect rather than emotions; cigarette smoking is declining in most developing countries as the health issues become better known.

Expectations

The strength of expectations is difficult to overestimate. So much of our behaviour is directly related to what we hope/expect/want to happen. Just imagine how house buyers would react if a major highway were to be built right through the neighbourhood they were looking at. Demand for the houses would drop like a paralysed parakeet due to the expectation of falling future property prices.[17]

Please note that expectation is one of many possible variables which will have an impact on demand. **Expectations** are noteworthy in the light of stock market downturns during 2008/9. It doesn't require a degree in rocket surgery to understand that when people expect certain things to happen in the future, it can well affect their demand in the present. As house buyers saw prices fall they held off on new house purchases – which led to a decrease in demand. This has led to a *'self-fulfilling prophecy'*. As housing prices fell, people expected this to continue. As consumers waited for a further fall in prices they actually held off on new housing purchases. This meant that demand fell… and that prices fell as an effect! This self-reinforcing negative feedback loop has proven immensely powerful for overall demand in the US economy since so much of household wealth is tied to property. As (perceived) wealth of home-owners declines, so too will demand for goods.[18]

Any time people expect things to happen within a foreseeable future time period, demand will be affected. If future prices are expected to rise, peoples' present demand might rise; if the future value of shares on the stock market is expected to rise, speculators' present demand for shares might rise; and if taxes on property are expected to fall in the future, present demand for property might rise. A recent example of speculatively driven demand is the price of guns in the US during 2009 due to expectations that President Obama was going to pass stricter legislation on arms sales.[19] An extreme example of the power of expectations occurred in Japan during the autumn of 2010 when the government passed laws for much higher taxes on cigarettes and Japanese smokers hoarded billions of dollars worth of cigarettes before the tax hike came into effect.[20] This actually had an effect on national income – perhaps adding as much as 1.4 percentage points[21] to national income.

---

15 I wonder if I can get away with including my favourite drink recipe. I know; "Dear colleague, Take a dark beer (not the English stuff – real beer) and pour into a large mug with two limes worth of juice and chili. Coat the rim of the mug with salt and chili. Now you have a delicious "*michelada cubana*"! *Salud*!

16 I love putting forth absolutely insane business ideas. One of these, which unfortunately came to me in an Energy Drink-induced vision blazed across the whiteboard while teaching and subsequently was worded out loud, was "chilli flavoured condoms". After the shrieks died down, one of the young ladies – with a look of complete seriousness – asked; "Em, Matt, flavoured on the inside or outside?" God I'm old.

17 This example of a 'self-fulfilling prophecy' is a notable element in speculative behaviour and possibly resulting in an upward sloping demand curve – which is a HL concept. I should

mention that the 2008-10 liquidity/credit/housing/stock market crisis was strongly linked to *speculative buying* on the housing market during the early 2000s.

18 This is known as the "wealth effect" and has a very powerful influence on the macro environment.

19 '*Gun sales decline as fear of curbs fades*', Business Week May 24 – 30

20 "Japan enjoys a nice, if fleeting, nicotine buzz" Business Week, Oct 4 – 10

21 It is important that you understand the difference between "increase in 1.4%" and "increase in 1.4 percentage points". If GDP (national income) is projected to grow by 3.5% and there is an increase of 1.4%, then this means a GDP increase of 3.549% (3.5 × 1.014). However, an increase of 1.4 *percentage points* means 3.5 + 1.4, i.e. 4.9%.

Population size and structure

An increase in population causing an increase in demand should be relatively plain. Yet there are also demand changes caused by a change in the *structure* of the population. For example, increased immigration will change the demand for certain types of food; age structure will change the demand for goods such as baby food or pensioners' homes; income and wealth re-distribution in a country will affect demand for income sensitive goods, and so on.

Derived demand

When demand for a good, say cars, increases then factor inputs such as steel will see an increase in demand. We say that the demand for steel is **derived** from the demand for cars. Derived means 'to be based on' or 'have its source in'. The concept is rather useful in understanding the interrelationship between goods that don't fall precisely into the category of complement goods. Derived demand is commonly associated with production, where firms using any number of products as inputs will affect demand for said inputs. Continuing with the example, car manufacturers use steel, rubber, plastic, glass…etc. A fall in the demand for cars will affect car manufacturers' demand for all of these inputs. During 2009/'10, rampant increase in Chinese demand for copper and iron from Australia and Brazil led to increased demand for Caterpillar mining machinery from the US manufacturer – a clear example of derived demand.[22]

One of my favourite examples of derived demand is currency, say the Namibian dollar, the demand for which is derived from the demand for tourism in Namibia; demand for Namibian exports; and the demand for investment in Namibia. Note that the causal flows can be reversed! Here in Mexico we receive a great deal of American tourists and when the US dollar appreciates (= increases in value compared to the Mexican peso) American tourists get more pesos for their dollar. Consequently, the US demand for hotels in holiday resorts such as Acapulco increases.

Putting the pieces together; oil and silicon

During 2007 and 2008, several economic commentators noticed how the increase in oil prices led to an increase in demand for silicon. Confused? It's really quite simple and deals with some of the issues outlined above, namely *substitutes* and *derived demand*.

- As the price of oil increased dramatically, households started to look to possible substitutes.

- Oil is used in many industries, one of which is energy. As energy costs for households rose, possible substitutes became more and more economically viable (= possible).

- One of the *substitutes* for oil in the field of energy is solar power. Thus, demand for solar panels increased drastically…

- …and since silicon is used in the manufacturing of solar panels, the ***derived demand*** for silicon rose and led to a tenfold price increase during 2007/2008.

On a personal note, I saw the writing on the wall early on in 2007 and bought shares in Vesta, a Danish company manufacturing wind power generators. Sometimes you win. Then oil prices fell drastically during 2009 and these shares bombed. Sometimes you lose.

MICROECONOMICS

---

22 "Stocks, copper rise, Treasuries drop..", Business Week 1 Sept 2010

HL extensions

In this HL section you will address a simple linear demand function and learn how the demand curve shifts and changes slope.

P ($)	Qd	Calculation
0	4,000	Qd = 4,000-20 × 0; **4,000**
50	3,000	Qd = 4,000-20 × 50; **3,000**
100	2,000	Qd = 4,000-20 × 100; **2,000**
150	1,000	Qd = 4,000-20 × 150; **1,000**
200	0	Qd = 4,000-20 × 200; **0**

Linear demand function (Qd = a – bP)

The demand curve plots out how a change in price causes a change in the quantity demanded. Since the demand curve is downward sloping, price and quantity are negatively correlated. We shall use the following form for our demand function:

$Qd = a – bP$

Qd: quantity demanded

a; is the *autonomous* level of demand (e.g. unrelated to price changes) – this is the Q-axis intercept

b: is the responsiveness of consumers to a change in price – this is the *slope*

P: price

We basically see that Qd is a function ('a result of a change in') the price of the good; $\Delta P \rightarrow \Delta Qd$. Assume that our D function is $Qd = 4,000 – 20P$

- At P(0) Qd will be 4000 units
- At P(200) Qd is zero units

It is conventional in Economics to plot Q (supply) on the horizontal and P (price) on the vertical axis. To meet with the mathematical conventions of graphing functions, we will quote each function as a 'P =' function as well as a 'Q =' function so that the slope may defined in the usual way as 'rise over run'.

Thus the demand function of this example:

$Qd = 4,000 – 20P$

may be transposed:

$20P + Qd = 4,000$ (by adding 20P to both sides)

$20P = 4,000 – Qd$ (by subtracting Qd from both sides)

$P = 200 – \frac{1}{20}Qd$ (by dividing both sides by 20)

Using conventional mathematical theory, the slope of this curve is $– \frac{1}{20}$ meaning that for every increase in Q of 1 unit, there is a reduction of price of $\frac{1}{20}$ (or an increase of 1,000 in Q results in a $50 reduction in price.

$Qd = 4000 – 20P$ every unit increase in P leads to a reduction of 20 units in Qd

OR

$P = 200 – \frac{1}{20}Qd$ every unit increase in Qd leads to a reduction of $\frac{1}{20}$ units in P

Slope of P relative to Q =
$–80 \div 1600 = –\frac{1}{20}$

Slope of Q relative to P =
$1600 \div –80 = –20$

+ 1600 units

–80 units
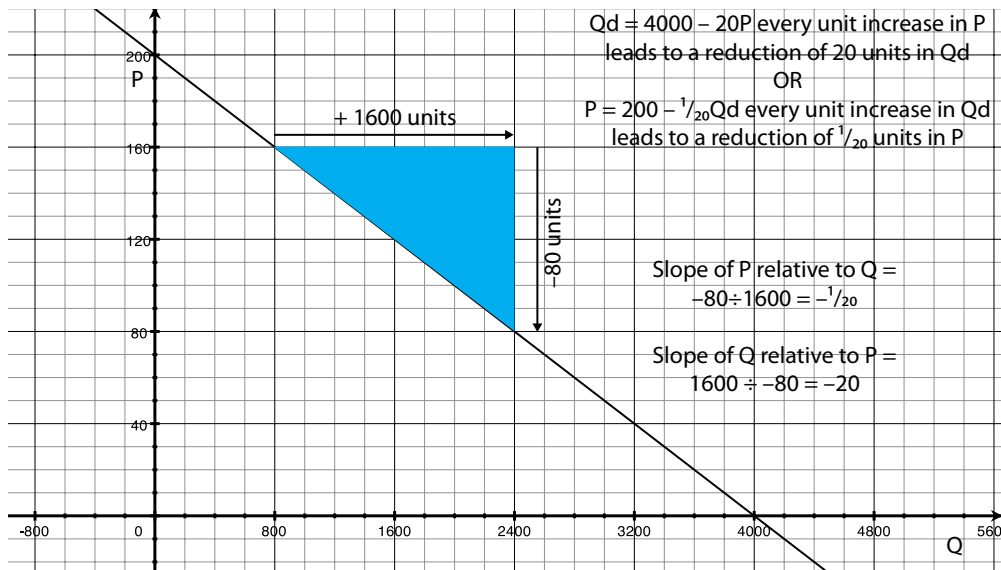
Figure 4.6

Shifting the demand curve

Recall that our demand function is Qd = a – bP. Any change in a non-price will shift the demand curve. Assuming that demand increases, this in fact means that quantity demanded increases at all price levels. This is a change in *autonomous* demand ('a').

Assume that the new D function is Qd = 5,000 – 20P…

or P = 250 – $^1/_{20}$Qd

- At P(0) Qd will be 5,000 units

- At P(250) Qd is zero units

P	Qd	Calculation
0	5000	Qd = 5,000-20 × 0; **5,000**
50	4000	Qd = 5,000-20 × 50; **4,000**
100	3000	Qd = 5,000-20 × 100; **3,000**
150	2000	Qd = 5,000-20 × 150; **2,000**
200	1000	Qd = 5,000-20 × 200; **1,000**

Note that every incremental increase in price of $50 *still* leads to a decrease in Qd of 1,000 units – e.g. the *slope* has not changed.
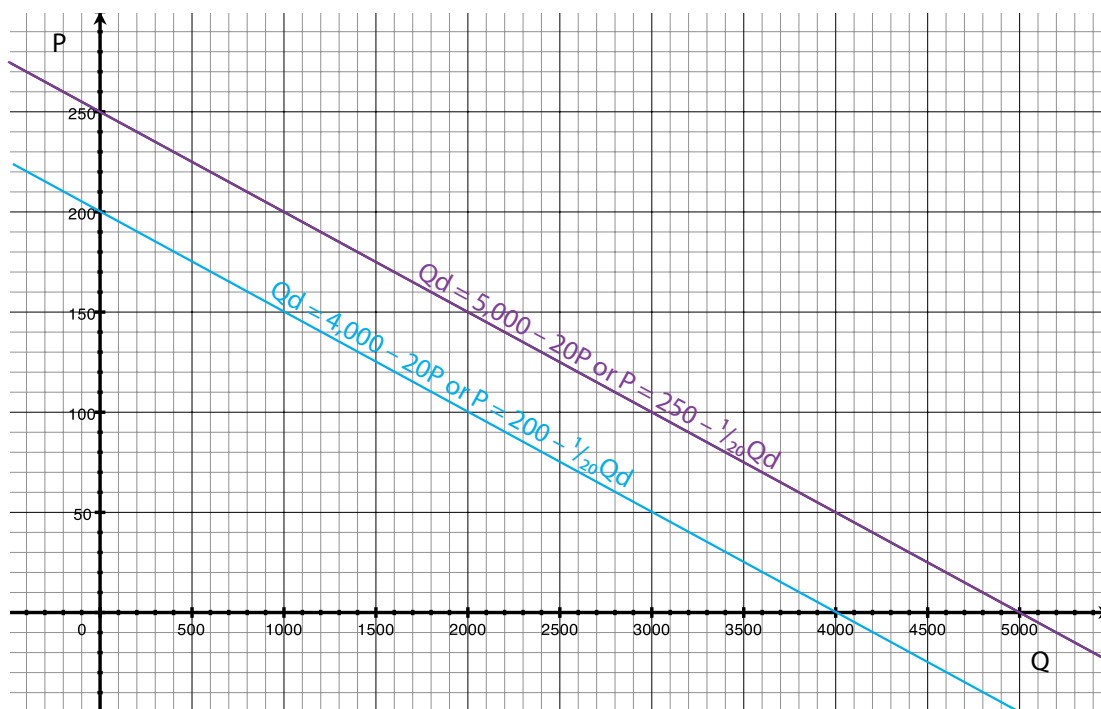
Figure 4.7

Changing the slope of the demand curve

So, our original demand function is Qd = a – bP which is Qd = **4,000** – **20P**. If the slope changes from 20 to 15, e.g. **Qd = 4,000 – 15P** …

- o The slope becomes steeper

- o The D-curve will intercept the P-axis at….[23]

Changing the slope from 20 to 15

P	Qd	Calculation
0	4000	Qd = 4,000-15 × 0; **4,000**
50	3250	Qd = 4,000-15 × 50; **3,250**
100	2500	Qd = 4,000-15 × 100; **2,500**
150	1750	Qd = 4,000-15 × 150; **1,750**
200	1000	Qd = 4,000-15 × 200; **1,000**

---

23   Stick in Qd = 0 into the formula; Qd = 4,000 – 15P → 0 = 4,000 – 15P → 4,000/15 = **266.67**

If *demand de*creases and the *slope* also decreases…

- The new function is **Qd = 3,000 – 30P** (P = 100–$^1/_{30}$Qd)

Any change in a non-P variable can affect both 'a' and 'b' in the D function

- A change in '**a**' means that demand has changed … (i.e. a change in a non-P variable!)

- And a change in '**b**' means a change in the responsiveness to a change in price (as in slope)

Changing the slope from 20 to 30 and decreasing demand

P	Qd	Calculation
0	3000	Qd = 3,000-30 × 0; **3,000**
50	1500	Qd = 3,000-30 × 50; **1,500**
100	0	Qd = 3,000-30 × 100; **0**
150	-1500	Qd = 3,000-30 × 150; **-1,500**
200	-3000	Qd = 3,000-30 × 200; **-3,000**

Any change in a non-P variable can affect both 'a' and 'b' in the D function

- A change in '**a**' means that *demand* has changed … (i.e. a change in a non-P variable!)
- And a change in '**b**' means a change in the *slope* of the demand curve (this is related to the *sensitivity* of the good in terms of a change in price – see price elasticity of demand in Chapter 9)
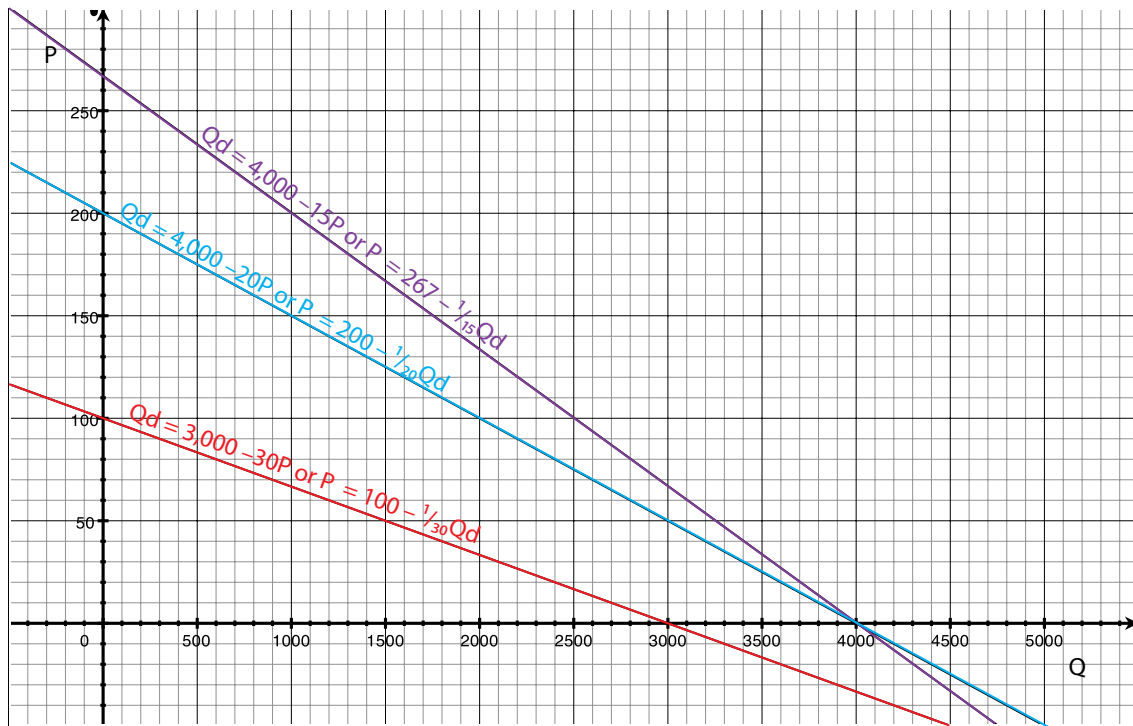
Figure 4.8

POP QUIZ 4.1

Note that calculations such as you will do in supply and demand will return several times in later sections.

1. Assume a demand function of **Qd = 200 – 2P**. This tells us that:

    i. The Q-intercept, 'a', is 200 units

    ii. The P-intercept is 'a' / 'b'; 200/2 = $100

    iii. The slope is 2; for each $1 increase in price, Qd decreases by 2 units

2. Make a simple table showing the Qd when P=200, 150, 100, 50 and 0.

3. Illustrate these values (pairs) in a diagram, e.g. draw a demand curve.

4. Draw a new demand curve based on the *price of a complement decreasing* – this leads to a 20% change in demand for our good but no change in slope.

5. Draw another demand curve in the same diagram showing that the price sensitivity of the good increases, e.g. the slope changes. For each $1 increase in price, the Qd decreases by 3. (Q-intercept remains unchanged.)

Summary & revision
Demand

1. Markets are a very efficient way of satisfying wants/needs in society and allocating resources via the price mechanism.

2. Four main market types are identified:

   a. **Perfectly competitive markets** – many small firms producing homogeneous goods and low barriers to entry for other firms
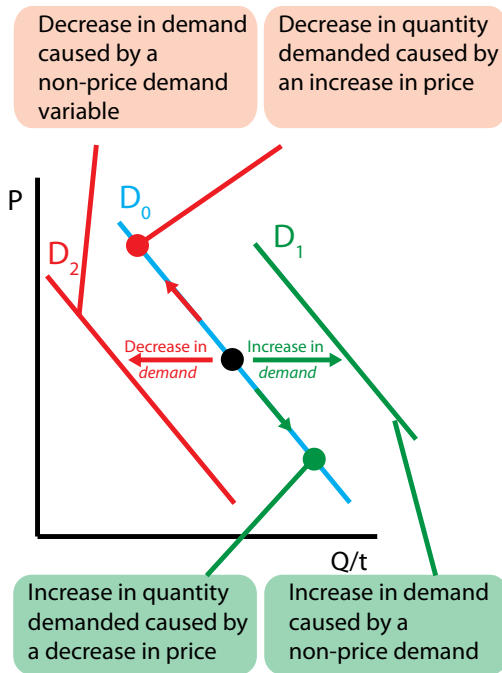
   b. **Oligopoly markets** – a few large firms dominate the market where entry barriers are high

   c. **Monopolistic competition** – many firms compete on a market with relatively low barriers to entry for differentiated (branded) goods

   d. **Monopoly** – one firm controls the market and barriers to entry are very high

3. **Law of demand**: an increase in the price of a good – *ceteris paribus* – will lead to a decrease in the quantity demanded

4. *Ceteris paribus*: all else equal, nothing else changes

5. **Effective demand**: the quantity of a good consumers are willing and able to purchase at different prices during a given time period

6. **Individual demand**: the willingness/ability of an individual consumer/firm to buy goods at different prices during a period of time

7. **Market demand**: sum of all individual demand

8. Downward sloping demand curve has two causes:

   a. Income effect – a decrease in the price of a good means real income rises for consumers who then can buy more of the good

   b. Substitution effect – a decrease in the price of a good (*ceteris paribus*) means that the relative price of substitutes has risen so consumers substitute a quantity of other goods for the lower priced alternative

9. A change in price leads to a movement along the demand curve

10. A change in a non-price variable affecting demand leads to a shift in the demand curve

Decrease in demand caused by a non-price demand variable

Decrease in quantity demanded caused by an increase in price

Increase in quantity demanded caused by a decrease in price

Increase in demand caused by a non-price demand

NOTE:

This and several subsequent chapters refer to plotting graphs and finding their slope. This topic is covered in more detail in a free downloadable set of resources. To get these, go to the author's website http://www.goodbadecon.com/

The resource can be found under 'Slope'.

5. Markets and supply

Key concepts:

- Law of supply and correlation between price and quantity supplied
- How individual firms' supply comprises market supply
- Introduction to basic market model; upward sloping supply curve
- Non-price determinants of supply
- Movement along vs. shift of supply curve

HL extensions:

- Linear supply function; $Qs = c + dP$
- Shifting the supply curve, e.g. a change in "c"
- A change in the slope of the supply curve , e.g. a change in "d"

MICROECONOMICS

Law of supply

One of the hippie credos of my troubled childhood was the T-shirt legend, 'What if they gave a war and nobody came?' This is like saying that a good has been provided but there is no demand for it whatsoever. There is no such thing; in order to be able to refer to a market, there have to be both willing buyers and willing sellers. One might actually say that demand creates supply.

The supply curve operates along the exact same parameters as demand. A *change in price* will – *ceteris paribus* – result in a *change in the quantity supplied*. As in the pattern of demand, there is clear correlation as long as other influencing factors remain the same. This is the **law of supply**, and differs from demand in having *positive correlation* – when the price rises, *ceteris paribus*, the quantity supplied increases and vice versa.

As so many issues in economics, this is intuitively obvious. An increase in the market price (for whatever reason) increases the propensity of suppliers to put the good on the market. There are a few simple reasons for this – all of which deal (once again) with the **willingness** and **ability** of suppliers to put goods on the market. Just as was done in the section on the demand curve, it is necessary to clarify that supply is not the amount that suppliers would *like* to supply, but rather the quantity they *intend* to sell (= are willing and able to sell) during a given time period.

> **Definition: 'Law of supply'**
>
> "*Ceteris paribus*, a rise in the price for good X will result in an increase in the quantity supplied of good X. A fall in the price of X will result in a decrease of quantity supplied."
>
> The law of supply states that as long as all other variables (the cost, availability and quality of factors of production) remain unchanged, then a rise in the price of a good will lead to an increase in the quantity supplied.
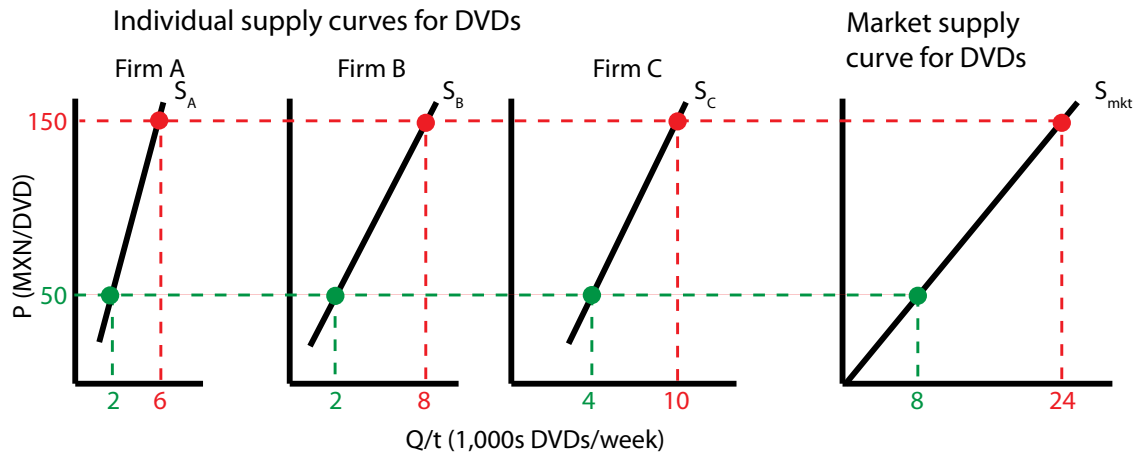

Figure 5.1: Unwilling seller!

Figure 5.2: Individual supply and market supply for DVDs in Naucalpan

Individual firms and the market supply curve

Just as market demand is derived (= developed) from the individual demand curves, so too is the market supply curve. Continuing with the simple example of a market for DVDs, let us assume there are three firms on the market selling DVDs.

In Figure 5.2, each of the firms has an individual supply curve. Firm A is able and willing to put 2,000 DVDs on the market at a price of MXN50 and 6,000 at a price of MXN150. The supply curves for Firm B and C show, respectively, 2,000 DVDs and 10,000 at these two prices. Summing up horizontally, the market supply curve shows that firms on the market will supply a total of 8,000 DVDs at a price of MXN50 and 24,000 at a price of MXN150.

Upward sloping supply curve

In Figure 5.3, if the price of DVDs rises from MXN100 to MXN150, the *quantity supplied increases* from 16,000 DVDs to 24,000 per week. This is in accordance with the law of supply; a higher price creates an incentive (the possibility of higher profit) for producers/suppliers to put more of the good on the market. We once again use the *ceteris paribus* assumption when drawing the supply curve. However, it is not incomes, preferences and other goods' prices that are kept constant (as we did in drawing the demand curve) but the **price, availability** and **quality of production factors** which we assume to be constant. (There are also various forms of government intervention which affect supply. More later.)

An increase in the price, *ceteris paribus*, from MXN100 to MXN150 causes a movement along the supply curve from A to C. This is termed an **increase in quantity supplied.**

A decrease in the price, *ceteris paribus*, from MXN100 to MXN50 causes a movement along the supply curve from A to B. This is termed a **decrease in quantity supplied**.
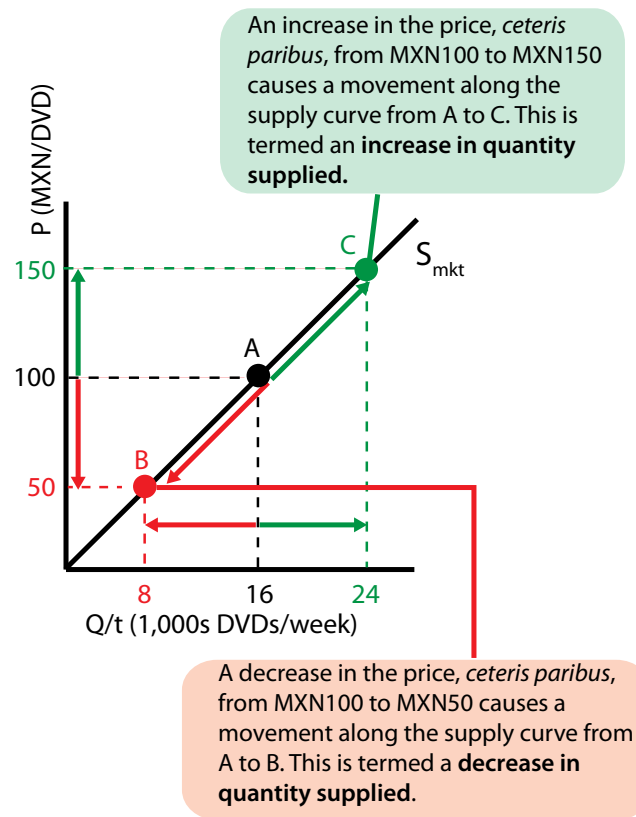
Figure 5.3: Market and upward sloping supply curve

Let's look more closely at the issue of the upward slope of the supply curve for DVDs in Naucalpan, Mexico. It is almost intuitive that higher prices would cause existing stores to increase available stocks of DVDs. There are two reasons for this **positive correlation**.

- The first is that a higher ticket price might add to firms' *revenue* (price times quantity), and perhaps also additional to *profit* (total revenue minus total costs – more in later chapters). Thus, if the market price rises

suppliers will have an **incentive** to put more of the good on the market. In the case of DVDs, sales outlets (suppliers) will order more stocks of DVDs from regional wholesalers and put more DVDs on their shelves.

- The other reason why an additional quantity is supplied at a higher price deals with the **costs** of producers. (A key issue here are *rising marginal costs*, a HL concept in Chapter 23, Theory of the Firm.) It is easy to understand that different suppliers have different cost levels. If the market price is MXN100 and the cost to a supplier in getting hold of an additional DVD to sell is MXN101…then basically the supplier would lose MXN1 for an additional DVD and it would not be provided. Now, if the market price rises to MXN102, then the supplier would not only be *willing* but *able* to provide the additional DVD.

DANGER MINES WARNING !

Law of Supply

Important note: 'Supply' is NOT the same as 'quantity in existence'!!! It is most important that you understand at an early stage that having 200 tonnes of cheese in a storage room is not the same as 200 tonnes being on the market. You see, if the producer of the cheese in question is simply storing the cheese then it is not actually offered on the market. It is not part of supply. It will become part of supply if the market price of this cheese increases to the level where the producer decides that it is not worth storing anymore and offers it for sale.

Thus: oil in an oil-field is NOT part of supply; oil in a cistern which is offered to a petroleum company IS part of supply. Schools of tuna fish are NOT part of supply; when tins of tuna held in a warehouse are offered to the supermarkets they ARE part of supply.

Movement along vs. shift of supply curve

In Figure 5.3 above, assume that the initial price is MXN100 and thus that *quantity supplied* per week is 16,000 DVDs. If the price increases to MXN150, then the quantity supplied increases from 16,000 DVDs per week to 24,000. This is shown by a *movement along* the supply curve, from **point A to C**,

which means, using correct economic terminology, that '…the *quantity supplied increases…*.' Accordingly, a decrease in the price from MXN100 to MXN50 is a movement along the supply curve from point A to B – the quantity supplied has decreased from 16,000 DVDs per week to 8,000.

Shifts in the supply curve

As will be done in some depth further on, any change in the price, availability, quantity and quality of factors of production will cause a change in supply – e.g. a shift in the supply curve. Figure 5.4 below shows that any change in a *non-price variable* off supply causes a shift in the supply curve.
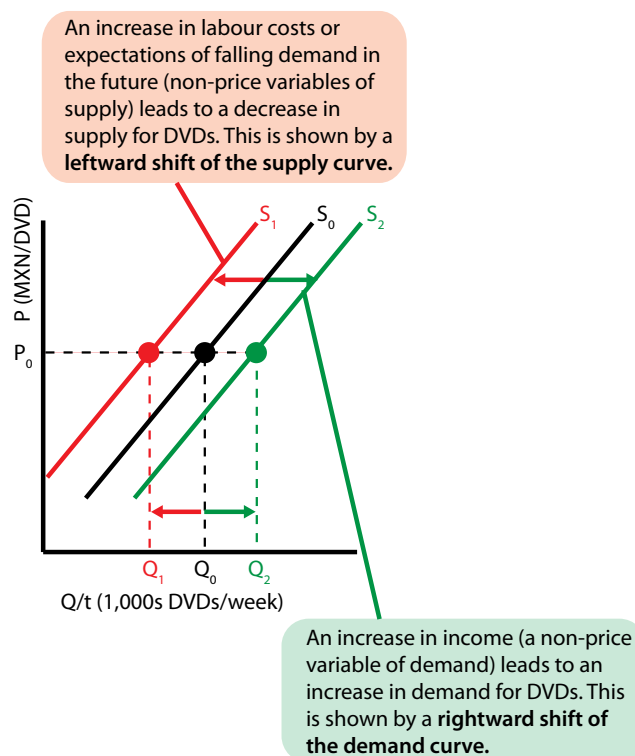
An increase in labour costs or expectations of falling demand in the future (non-price variables of supply) leads to a decrease in supply for DVDs. This is shown by a **leftward shift of the supply curve.**

An increase in income (a non-price variable of demand) leads to an increase in demand for DVDs. This is shown by a **rightward shift of the demand curve.**

Figure 5.4: Shifting the supply curve for DVDs

Non-price determinants of supply

Just as in demand, there are "other variables" affecting supply within the *ceteris paribus* assumption. These are **non-price determinants of supply**. Any change in a non-price determinant of supply means that the pattern of supply changes – shown by a *shift* of the supply curve. Profit incentive and covering increasing (marginal) costs explain why the supply curve is upward-sloping and (just as in the pattern of demand) there are a number of non-price determinants of supply, all of which will *shift the supply curve.*

> **Definition: 'Non-price determinants of supply'**
> Any variable that changes the pattern of supply other than price is a non-price determinant of supply. Changes in the price, availability, quality and quantity of factors of production will all have an effect on the supply of goods.

The non-price determinants of supply can be divided into four main categories:

1. Changes in relevant **market factors:** these are costs of factors of production; price of related goods (producer substitutes); firms' expectations; and market entry/exit

2. Changes in availability/**scarcity of factors** of production: these include reserves of natural resources; weather and climate changes; natural disasters

3. Changes in the **quality/quantity of factors** of production: factor inputs improve over time due to improved production methods; technological advances, and advances in materials

4. **Market intervention** (non-market variables) by government such as taxes and subsidies.

1. A change in relevant market factors

Cost of factors of production: Any increase in costs to suppliers will mean that the cost of producing goods will increase. This lowers the ability and willingness of suppliers to put the goods on the market (since they cannot influence the market price) and producers will respond by *decreasing* output. Decreased raw material prices, lower labour costs and lower rents would all cause production costs to decrease and supply to increase. Thus, if wages or salaries rise, then the workers used in production become more costly whereupon producers will decrease output – e.g. the supply curve shifts left. Substituting 'Swiss watches' for 'DVDs' in Figure 5.4 illustrates how an increase in labour costs for Swiss watch manufacturers would cause a *shift* in the supply curve to the left, from S_0 to S_1, i.e. fewer watches are supplied at P_0 and at *all* prices.[1]

---

1   I did not use labour costs as an example randomly! Raw material costs – even such precious metals as gold and platinum – represent a fraction of the final value of a good Swiss watch. Perhaps the most complicated watch in the world, "Calibre 89" made by Patek Philippe for its 150th anniversary in 1989, contains 1728 parts, and had four people working on it for nine years (!) measured from initial research/design to final completion. When I get the required USD6 million together from book sales I'm going to get one.

Price of related goods: Producers will have any number of possible *producer substitutes* (not to be confused with consumer substitutes!) such as DVDs and Blu-Ray discs. An increase in demand/price for Blu-Ray discs might cause suppliers to reallocate (= shift) resources from the production of DVDs to Blu-Ray discs, causing the supply of (the producer substitute) DVDs to decrease from S_0 to S_1 as shown in Figure 5.4.

Another possibility is that producers have goods that are in *joint supply*. An example would be gold and copper which are often found in the same geological veins and pockets. An increase in the price of gold would lead to an increase in the quantity supplied of gold (a movement along the supply curve for gold) which in turn would increase the supply of copper, illustrated by the rightward shift in supply (S_0 to S_2) in Figure 5.4, In a similar vein (pun intended), a decrease in the price of beef would lead to a *decrease in the quantity supplied of beef* (movement along the supply curve for beef) and therefore also a *decrease in the supply of leather* (shift left of the supply curve for leather).

* **The expectations of firms:** If firms expect a surge in demand, such as seasonal demand for tourism, they might actually increase supply of, say, rental cars in order to build a sizable stock for the *coming* increase in demand. Figure 5.4 shows how anticipation of the tourist season in Cancun, Mexico, could lead suppliers of rental cars to increase supply from S_0 to S_2.

* **Market entry/exit:** If existing firms on a market are making a profit, newcomers might be attracted. As firms enter the market, total supply of goods will increase. This is often the case for new products, where the success of the iPhone resulted in a rapid increase in "copycat" (= similar, copied) products. In the same way, if an industry becomes less profitable then firms will leave the market, decreasing supply, for example the decline in the Swedish ship-building industry in the 1980s caused by far more competitive ship-builders in Korea. (S_0 to S_1 in Figure 5.4).

2. A change in the availability/scarcity of factors of production

* **Availability and scarcity of factors:** Any and all factors used in production are subject to relative scarcity. Should factors become more available or abundant, then the same market laws will apply to them, i.e. scarcity will *raise the price* of any given factor of production. Studies show that during the end of

the decade, teachers in Sweden will be retiring at a rate far beyond the rate of replenishment via teacher colleges. This will decrease the supply of teachers – and hopefully raise the salaries of those of us who remain in the profession[2]. As for tangible (= physical) goods, suppose that the existing diamond mines in Botswana, Namibia, South Africa and Russia start to peter out (= diminish). This would mean that there would be less diamonds available no matter what we are willing to pay for them. While there is a nifty Swedish firm selling jewellery made from elk excrement, I don't envision this being a viable substitute in the context of weddings.

- *Force majeure*, **unexpected events:** Disruptions such as earthquakes, fires, floods and other natural disasters can have significant effects on the supply of goods – not only agricultural goods but all goods needing transport. Look at the frequent disruptions to oil supply; every time there is a terrorist attack (Iraq; blown up pipelines), flood (Mexico; workers can't get to the off-shore wells) or hurricane (Texas; destroyed oil refineries) there is a resulting decrease in the supply of oil and petroleum products.

Hurricane Frances gathers force in August 2004. The photo was taken from the International Space Station.

In the same mode, an increase in the supply of factors will both increase availability and lower the costs for producers. Just imagine how the discovery of new oil fields would help suppliers all around the world by lowering the dependency on the oil cartel, OPEC. Another thing mentionable in this context is the effect of weather conditions on agricultural output. During 2008 tens of thousands of people in Ethiopia once again faced starvation according to the FAO[3] due to lower than normal rains and poor harvests.

3) A change in the quality/efficiency of factors of production

- **The quality of factors of production:** Anything which enables a producer/supplier to put more on the market without increasing costs means that supply increases, i.e. the supply curve shifts to the right. Better production methods and/or a more educated and well-trained labour force would increase output during any given time period. This is an increase in the efficiency in use of factors – enabling the subsequent increase in productivity. *Production techniques* (division of labour for example), *advancing technology* (computer assisted design and computer assisted manufacturing – CADCAM to name but one), new and *improved materials* in production (any number of ceramics), and of course anything dealing with increasing *knowledge*, research and development (R & D), education, training….etc.

- **Technology:** Better tools, production processes, materials, computer assisted design and computer assisted manufacturing – CADCAM …etc, are all elements of advancing technology which increase output per unit of time, which is the same as increasing supply. For example, it now takes General Motors in North American an average of 24.4 hours to assemble a vehicle – 6.4% shorter time than the year before.[4]

4. Market intervention – taxes and subsidies

In addition to the forces arising from within the market itself, there are also forces which can be said to be imposed from without, i.e. market intervention by government. Two common ways for governments to intervene are by imposing an **indirect tax**[5] on goods or subsidising goods. An indirect tax – an expenditure tax – such as value-added tax (VAT) on goods sold is a percentage increase of the sales price of the good and will in effect be an increase in costs for producers, thus decreasing supply. A subsidy is the opposite, a payment to producers (often

---

3   Food and agricultural organization of the United Nations.

4   *US carmakers' efficiency rises…*, Business Week 30 June 2003

5   Indirect taxes are taxes which indirectly go to government, since VAT is paid to the suppliers who then in turn transfer the taxes to government. A direct tax is income and profit (corporate) tax, since this goes directly to government. We get shafted either way.

2   Maybe I should put in an advance order on a new Rolex.

per unit of output, for example per tonne of wheat) which lowers producers' costs (HL: marginal costs) and also acts as an incentive to produce more. This increases supply. (More in Chapters 13 and 14.)

Definition: 'Indirect tax" and 'subsidy'

An **indirect tax** such as an expenditure tax is added to the price of the good and causes suppliers to increase the price at all levels of supply. The supply curve shifts left.

A **subsidy** is a payment or money grant to suppliers. This works as an incentive to produce more and also lowers (marginal) costs of production. The supply curve shifts right.

OUTSIDE THE BOX

'Mermaidomics' revisited – 'backward bending' S-curve?!

During the speculative oil price shock of 2007/'08, several economic commentators observed that oil suppliers in fact had an incentive to decrease output of oil as the price rose. This seemingly contradicts economic theory, yet in fact it can be explained quite logically.
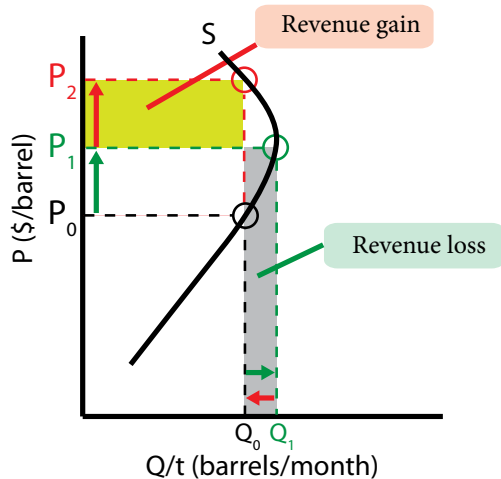

Figure: 5.5 Supply of oil

The price of oil is at P_0 and rises to P_1; the quantity supplied of oil increases to Q_1. Now, assume that:

1. Oil production starts to reach maximum (short run) output potential, and

2. Oil producers have a level of target revenue, i.e. an aimed-for level of oil revenue.

Two forces kick in which in fact might cause oil suppliers to decrease the quantity supplied; expectations of higher future oil prices and limited investment alternatives.

Expectations: Assume that suppliers are satisfied with the oil revenues at P_1. This is the area given by $P_1 \times Q_1$. A rise in price to P_2 means that suppliers can in fact decrease quantity supplied on the oil market (Q_2) while retaining the target revenue. (The loss of revenue – grey area – is equal to the gain in revenue shown by the green area.) If suppliers expect oil prices to remain high or continue upwards, they have an incentive to hold off on increasing capacity and supplies – they can sell less and still make at least the same revenue in the future.

Investment alternatives: Oil is a finite resource. Any oil left in the ground is a form of investment. Basically, by leaving the oil in the ground and – in line with the expectations-based argument above – betting on permanently higher oil prices, producers have "invested" in future oil revenues. As suppliers are reaching maximum short run output levels at P1, increasing costs of extraction above P1 simply might lead suppliers to hold off on production increases.

With help from http://web.mit.edu/krugman/www/opec.html

Summary of supply shifts

By now you should be accustomed to the methodology I use in explaining concepts. I try to follow a 'define – exemplify – context' formula, which I realise is both time consuming and frequently tiresome. Therefore, I shall simplify the iteration of changes in supply by using the nine additional examples used previously in a table below and labelling them accordingly in Figure 5.6.

1.   A change in the relevant market factors

Example:	Effect:	Supply will:	In fig: 5.6
a) wages for road workers fall	→ cost of making roads falls	supply curve for roads shifts right	S_0 to S_2
b) price of steel rises	→ cost of making cars rises	supply curve for cars shifts left	S_0 to S_1
c) price of flour falls	→ cost of making bread falls	supply curve for bread shifts right	S_0 to S_2

2.   A change in the availability/scarcity of factors of production

Example:	Effect:	Supply will:	In fig:5.6
d) fewer study medicine	→ fewer doctors in hospitals	supply curve for health care shifts left	S_0 to S_1
e) pilots go on strike	→ fewer flights available	supply curve for air travel shifts left	S_0 to S_1
f) new copper deposits found	→ more copper available	supply curve for electrical wire shifts right	S_0 to S_2

3.   A change in the quality/efficiency of factors of production

Example:	Effect:	Supply will:	In fig:5.6
g) increased use of robots	→ industrial production is faster	industrial supply curve shifts right	S_0 to S_2
h) better basic education	→ all labourers more efficient	supply curve for most goods shifts right	S_0 to S_2
i) IT revolution	→ most production more efficient	supply curve for most goods shifts right	S_0 to S_2

4.   Non-market variables, i.e. intervention

| Example:	Effect:	Supply will:	In fig:5.6
j) increased restrictions on cigarette sales	→ fewer sales outlets for cigarettes	supply curve shifts to the left	S_0 to S_1
k) taxes on gasoline are increased	→ firms must pay a portion of sales revenue to government	supply curve shifts to the left	S_0 to S_1
l) a subsidy is granted for the production of milk	→ milk producers have an incentive to produce more milk and production costs decrease	supply curve shifts to the right	S_0 to S_2

Figure 5.6

MICROECONOMICS
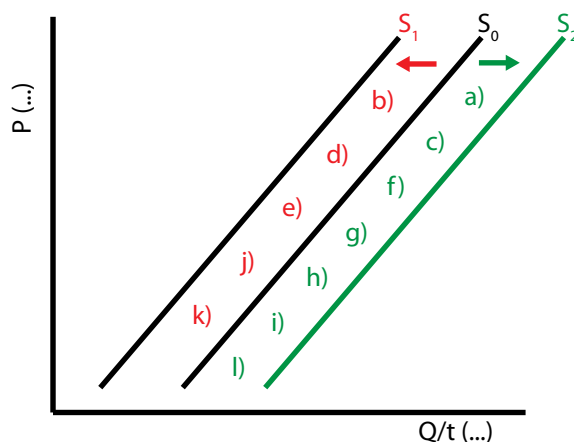
What we now need to do is put the supply and demand curves together in the same diagram, in order to be able to analyse market behaviour. It bears repeating: nothing done so far has equipped you with the necessary tools to do a market analysis! For this you need supply and demand working together, not the separate (purely explanatory!) graphs done thus far. Let's put it all together people.

Putting the pieces together

Wage negotiations

A few years ago I had a horrendous row with my (then) boss.[A] What the issue boiled down to was my unwillingness to accept more teaching hours and more students without a substantial rise in my salary.[B] Being me, I used a great deal of harsh language – most of it in print – and started to pack my bags and empty my computer. I was simply not prepared to give up any more of my increasingly valuable (i.e. scarce!) free time without increased remuneration (=payment). One could say that my opportunity costs of teaching additional classes were increasing and that I wanted my income to match this increased cost. This would render Figure 5.6 below. (We are assuming that my supply curve would be 'normal', i.e. I would supply more labour at a higher salary.)
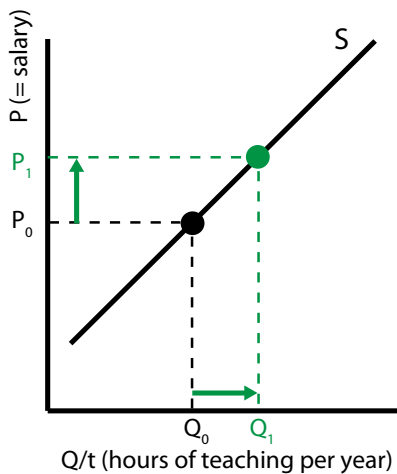

Figure 5.7: Matt's supply of teaching hours

Of course it is possible! My school could entice me to supply more teaching hours without paying me more – i.e. change my overall willingness, ability and propensity to supply teaching hours. Any suggestions? Perhaps giving me my own office –

preferably equipped with a hot computer, cold bar, colour TV, sofa and view. Or relieving me of the need to attend mind-numbing teachers meetings in the study hall?![C] Or extending all my vacations?! Or giving me complete and utter power over all the language teachers at school?!

You get the drift. This would mean more hours taught at all price levels. This is shown in Figure 5.8. If, for example, I accepted a better office rather than an increase in pay, I would provide additional teaching service at the same price – increasing my supply from S_0 to S_1. As the price hasn't changed, and the amount of teaching hours supplied has increased, supply has increased I have increased my willingness to supply teaching hours at all prices.
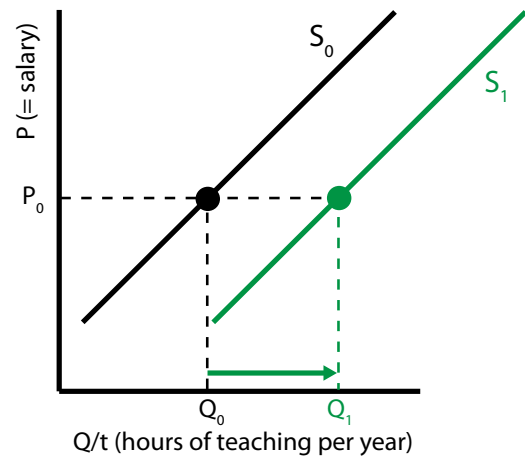
Figure 5.8: Matt's supply of teaching hours

A   I have in fact more ex-bosses than ex-wives.

B   A wage means "x dollars per hour", i.e. an amount of money per labour hour. A salary is not based on an hourly rate but is a fixed sum per month. Auto-workers get wages, teachers get screwed salaries.

C   No wait – I didn't attend those anyhow. Oh well.

HL Extensions

Linear supply function; Qs = c + dP

We now do the other hand – the upward sloping S-curve. Same type of formula but we use other letters so as not to confuse things with the D-function. Also, we will need to calculate equilibrium later on so we will need four different letters for the parameters – two for the demand function and two for the supply function.

$Qs = c + dP$

Qs: quantity supplied (at a given price)

c; autonomous level of supply (e.g. unrelated to price changes)

d: is the responsiveness of quantity supplied to a change in price – this is the *slope*

P: price

We basically see that Qs is a function ('a result of a change in') the price of the good; $\Delta\uparrow P \rightarrow \Delta\uparrow Qs$. Assume that our (linear) supply function is $Qs = -2,000 + 40P$ – at P=0 Qs will be -2,000

units – which is silly in realistic terms but simply shows that the supply curve will have a positive value on the P-axis, e.g. suppliers have a 'minimum price' before they supply any units at all on the market.

The P-intercept is given by $\frac{c}{d}$; $\frac{2,000}{40} = 50$.

Thus suppliers demand a minimum of $50 to put any units on the market (At P=$50 Qs is zero units.) The constant 'd' is 40 – this tells us that for every increase in price of $1, suppliers will be able and willing to put an additional 40 units on the market. Figure 5.9 illustrates the initial supply curve.

Original: Qs = -2,000 + 40P

P ($)	Qs	Calculation
0	-2,000	Qs = -2,000 + 40 × 0; -2,000
25	-1,000	Qs = -2,000 + 40 × 25; -1,000
50	0	Qs = -2,000 + 40 × 50; 0
75	1,000	Qs = -2,000 + 40 × 75; 1,000
100	2,000	Qs = -2,000 + 40 × 100; 2,000

The equation Qs = -2,000 + 40P transforms to $P = \frac{Qs + 2,000}{40}$ or $P = \frac{Qs}{40} + 50$. It is in this form that the graph can most easily be plotted with P on the vertical axis and Qs on the horizontal axis.
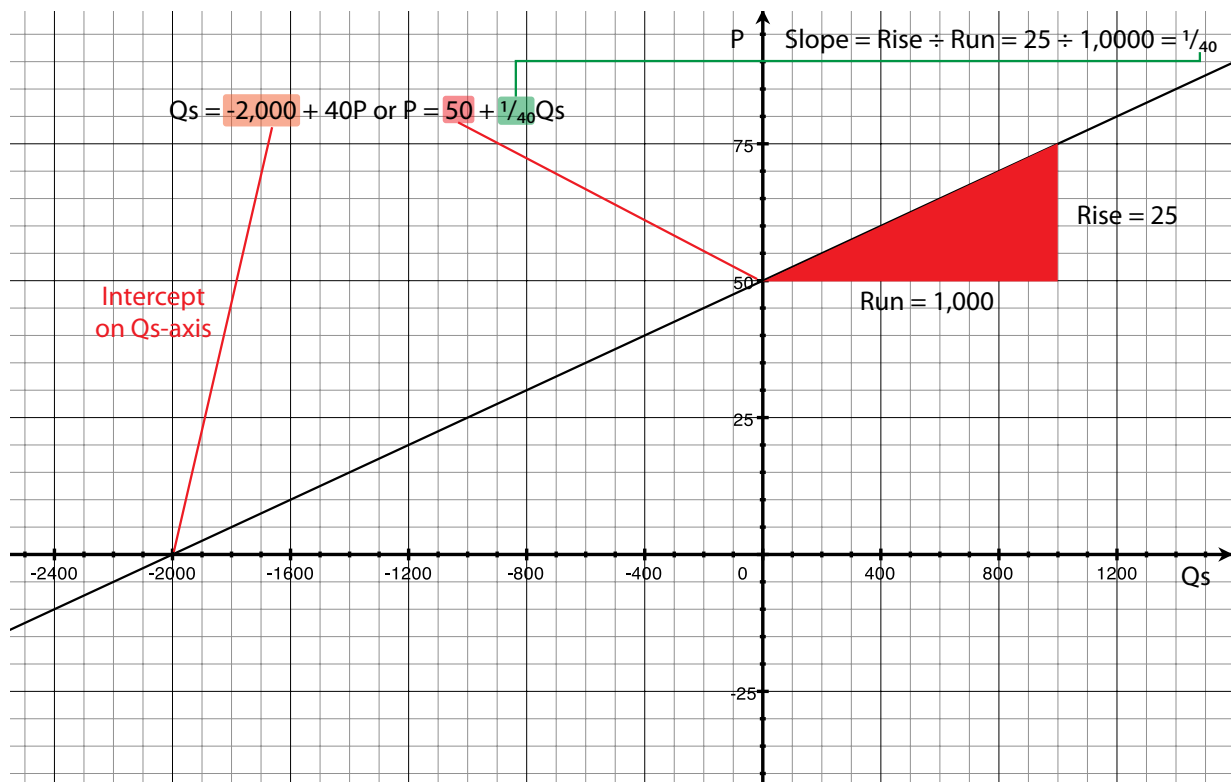

Figure 5.9 Supply curve for the function $P = \frac{Qs}{40} + 50$ (or Qs = -2,000 + 40P)

Shifting the supply curve, e.g. a change in c

Any change in c will result in a shift of the demand curve, shown in Figure 5.10:

- **Decrease**; if supply decreases by 1,000 units at each price level, then there is a leftward (parallel) shift of the supply curve and the Q-intercept is -3,000. If supply decreases by 50% at all prices

  o The P-intercept is $\frac{3,000}{40}\left(\frac{c}{d}\right) = \$75$

- **Increase**; if supply increases by 1,000 units at each price level, then there is a rightward (parallel) shift of the supply curve and the Q-intercept is -1,000.

  o The P-intercept is $\frac{1,000}{40}\left(\frac{c}{d}\right) = \$25$

Supply decrease: Qs = -3,000 + 40P

P ($)	Qs	Calculation
0	-3,000	Qs = -3,000 + 40 × 0; -3,000
25	-2,000	Qs = -3,000 + 40 × 25; -2,000
50	-1,000	Qs = -3,000 + 40 × 50; 1,000
75	0	Qs = -3,000 + 40 × 75; 0
100	1,000	Qs = -3,000 + 40 × 100; 1,000

Supply increase: Qs = -1,000 + 40P

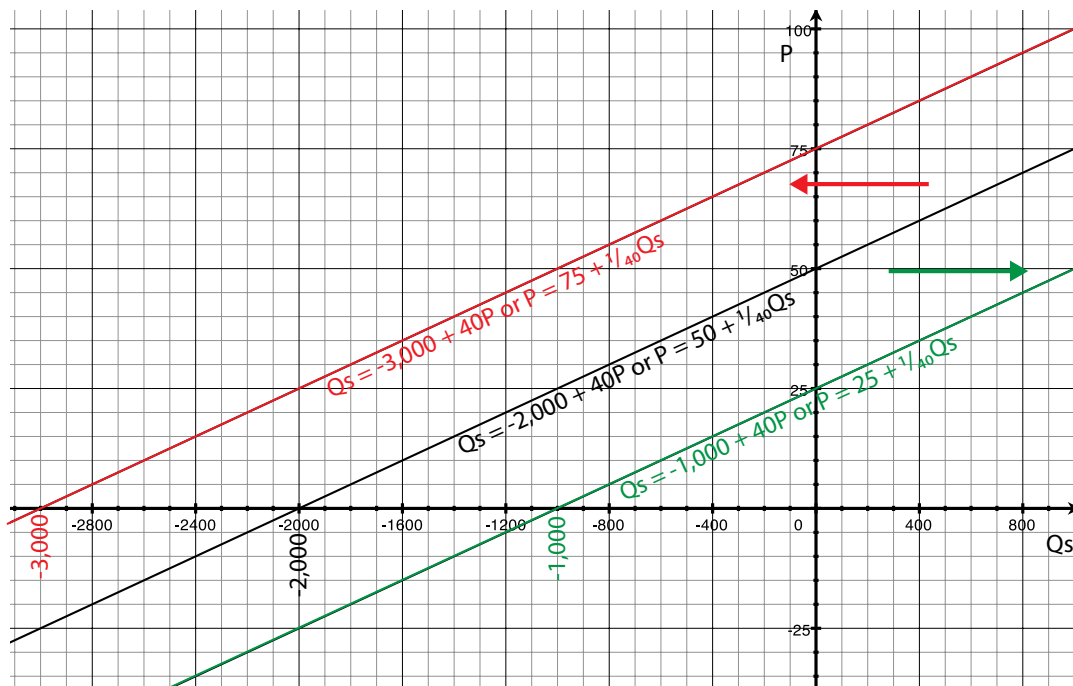| P ($)	Qs	Calculation
0	-1,000	Qs = -1,000 + 40 × 0; -1,000
25	0	Qs = -1,000 + 40 × 25; 0
50	1,000	Qs = -1,000 + 40 × 50; 1,000
75	2,000	Qs = -1,000 + 40 × 75; 2,000
100	3,000	Qs = -1,000 + 40 × 100; 3,000

Figure 5.10 Shifting the supply curve

A change in the slope of the supply curve , e.g. a change in d

A change in d will of course result in a change in the slope, Figure 5.11:

- Assume that the new supply function is **Qs = -2,000 + 50P** or P = $^1/_{50}$Qs + 40

  o Basically, this tells us that for any change in P of $1, the Qs will increase by 50 units

  o An increase in price of 25 → Δ↑Qs of 1,250

  o An increase in price of 50 → Δ↑Qs of 2,500

  o The Q-intercept is still -2,000 but the

  P-intercept is $\frac{2,000}{50}\left(\frac{c}{d}\right) = \$40$

- A supply function of **Qs = -2,000 + 20P** or P = $^1/_{20}$Qs + 100 increases the slope (from $^1/_{50}$ to $^1/_{20}$)

o A change in price of $1 causes an increase in Qs of 20 units

o The Q-intercept is still -2,000 but the

P-intercept is $\frac{2,000}{20} \left(\frac{c}{d}\right) = \$100$

Supply increase in slope: Qs = -2,000 + 20P

P ($)	Qs	Calculation
0	-2,000	Qs = -2,000 + 20 × 0; -2,000
25	-750	Qs = -2,000 + 20 × 25; -1,500
50	500	Qs = -2,000 + 20 × 50; -1,000
75	1,750	Qs = -2,000 + 20 × 75; -500
100	3,000	Qs = -2,000 + 20 × 100; 0

Supply decrease in slope: Qs = -2,000 + 50P

P ($)	Qs	Calculation
0	-2,000	Qs = -2,000 + 50 × 0; -2,000
25	-750	Qs = -2,000 + 50 × 25; -750
50	500	Qs = -2,000 + 50 × 50; 500
75	1,750	Qs = -2,000 + 50 × 75; 1,750
100	3,000	Qs = -2,000 + 50 × 100; 3,000

The shift (Δc) and changes in slope (Δd) are of course caused by changes in the all the usual non-price determinants of supply such as a change in the price/availability of raw materials and labour, market entry by firms, increases in productivity and market intervention such as taxes and subsidies.
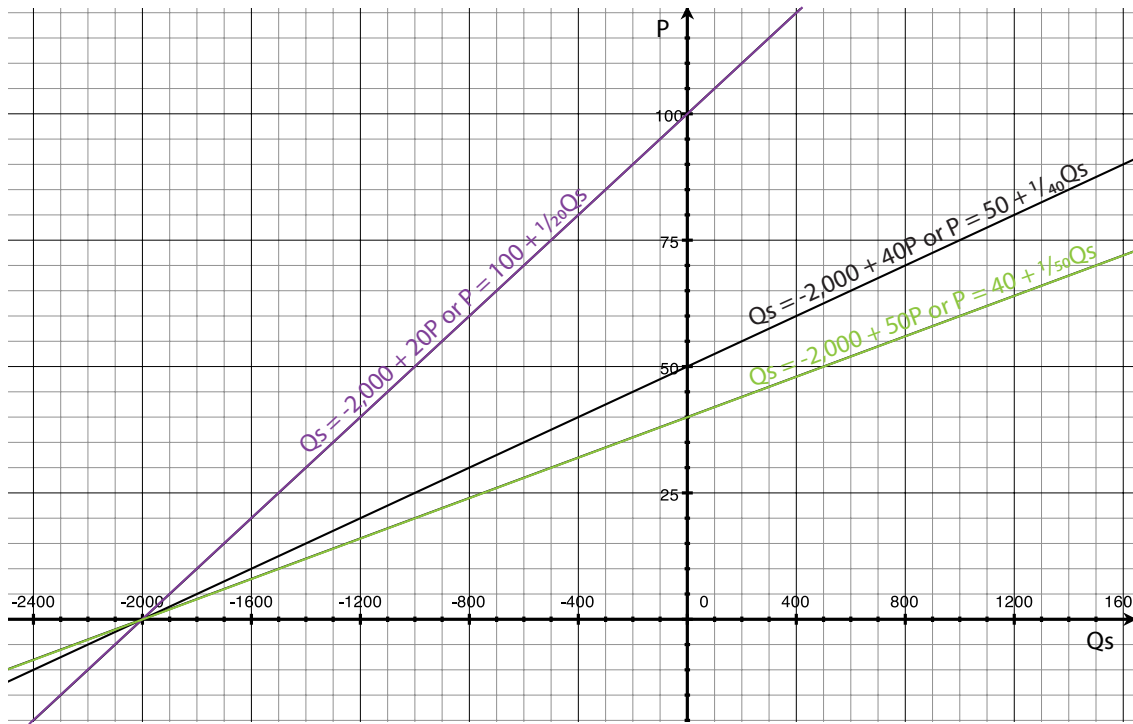
Figure 5.11 Changing the slope of the supply curve

POP QUIZ 5.1

1. Assume a supply function of **Qs = 200 + 10P**. This tells us that:

   i. This is an *inelastic supply curve* – e.g. the supply curve intercept on the Q-axis will be a positive value (at a price of zero quantity supplied is 200)

   ii. The P-intercept will be negative (solving for Q = 0 gives us a price of -20)

   iii. The slope is 10; for each $1 increase in price, Qs increases by 10 units

2. Make a simple table showing the Qs when P=0, $10, $20, $30 and $40

3. Illustrate these values (pairs) in a diagram, e.g. draw a supply curve.

4. Draw a new supply curve based on the *price of raw material decreasing* – this leads to a 20% change in supply for our good but no change in slope.

5. Draw another supply curve in the same diagram showing that the slope changes. For each $1 increase in price, the Qs increases by 8. (Q-intercept remains unchanged.)

b. Change in the price of producer substitutes.

c. Expectations of firms.

d. Market entry/exit by other firms.

e. Government intervention (taxes and subsidies).

Summary & revision

1. **Law of supply**: an increase in the price of a good – *ceteris paribus* – will lead to an increase in the quantity supplied

2. **Individual supply**: the willingness/ability of an individual firm to put goods on the market at different prices during a period of time

3. **Market supply**: sum of all individual supply curves

4. Upward sloping supply curve has two causes:

   a. Incentives – if the market price rises for a good suppliers will have an incentive to put more of the good on the market.

   b. Increasing costs – as costs increase for producers in putting additional units on the market, only higher market prices will induce them to increase the quantity supplied in that increasing (HL: marginal) costs may be covered.

5. A change in price leads to a movement along the supply curve .

6. A change in a non-price variable affecting supply leads to a shift in the demand curve, e.g.:

   a. Change in price/availability/quantity or quality of factors of production.

Decrease in *supply* caused by a non-price supply variable.

Increase in **quantity supplied** caused by an increase in price

Decrease in supply

Increase in supply

Decrease in **quantity supplied** caused by a decrease in price

Increase in **supply** caused by a non-price supply variable
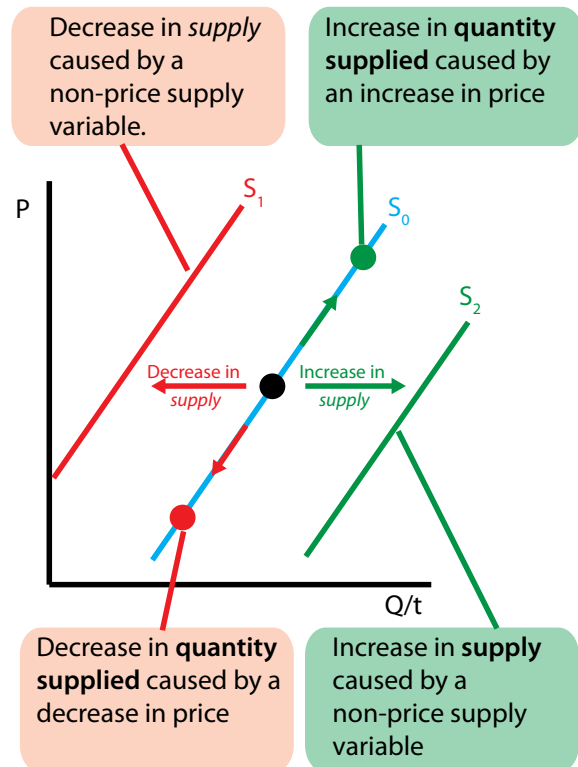
HL extensions

7. Linear supply function: $(Qs = c + dP)$.

8. Shifting the supply curve (Δc).

9. Changing the slope of the supply curve (Δd).

6. Market equilibrium

Key concepts:

- Excess demand and excess supply
- Equilibrium (market clearing)
- **Change in demand** (e.g. shift of demand curve) à change in quantity supplied (e.g. a movement along the supply curve)
- Tastes/preferences, price of substitutes, price of complements, advertising...
- **Change in supply** (e.g. shift of supply curve) à change in quantity demanded (e.g. a movement along the demand curve)
    - Market factors
    - Scarcity of factors of production
    - Quality/quantity of factors of production
    - Non-market variables, i.e. intervention

HL extensions:

- Equilibrium price and quantity – linear supply and demand functions
- Plotting linear supply and demand curves to identify equilibrium price and quantity
- Excess demand/supply in diagrams.

If it's not part of the solution...

...then it's part of the problem. The simple function of a market is to put both parties – customers and suppliers – in touch with each other. The rest follows: any given want of a potential buyer will result in search behaviour, and the opportunity costs of that search are lowest where a great many items may be seen and compared in the shortest amount of time. The lower opportunity cost is thus; the potential buyer can compare a great many similar goods in short order and perhaps bargain the way towards the highest satisfaction. This is far more efficient than gallivanting around the countryside in search of 20 different items. The market is simply a very efficient method of displaying the merchandise.

The same holds for providers of goods, as the demand is a signal to suppliers as to what goods should be provided and what the going price is. If the price is attractive and/or the demand is high then rest assured that the supplier will be back the next time with more goods for sale! Should there be too many similar goods for sale, then one can assume that a number of suppliers will put effort into finding other or better goods to sell – or other places to sell them. If the going market price is too low for a supplier to be able to sell his/her wares, then there is a strong incentive for the supplier to increase efficiency in order to lower costs – and thus compete in the market.

The market thus functions as a mechanism to fulfil both customers' and suppliers' wants. The market system also addresses the issue of excess and quality; any superfluous goods are simply not sold at the existing price. Market dynamics have thus solved the basic economic problem of what, how and for whom: consumers' demand decides what is to be produced; basic competition between suppliers addresses the issue of how to produce (as inefficient producers are forced to leave the market); and the price decides who gets the goods.

As long as we are all good comparison shoppers and firms operate from within competitive forces, the system will provide the 'right amount of goods' at the 'correct price'. In theory, that is. Unfortunately, as we shall see, for the tooth fairy to deliver, one has to knock out a few teeth.

Excess demand, excess supply and equilibrium

"I am like any other man. All I do is supply a demand." (Al Capone)

In class, I often bring up the open markets I've visited. I describe them in some detail in terms of the time spent putting shoddy goods out for sale. I still claim that many of them obey only half the law of supply/demand; if something 'sucks' there's always plenty of it. I have more or less observed this in virtually every country I've been to. My conclusion; too many people have too much free time on their hands.

In any case, market clearing – e.g. *market equilibrium* – is at the heart of the concept of market dynamics, which is based on how the interaction of supply and demand creates a *market price*. Let us continue with the market for DVD films.

Figure 6.1 Quantity (1,000s) of DVDs demanded/supplied per week

Price (MXN)	Quantity supplied	Quantity demanded	Excess supply
150 ·	24	8	16 (excess quantity supplied)
125	20	12	8 (excess quantity supplied)
100	16	16	Market clearing price!

Price (MXN)	Quantity supplied	Quantity demanded	Excess demand
75	12	20	8 (excess quantity demanded)
50	8	24	16 (excess quantity demanded)

The table in Figure 6.1 shows the market for DVDs. The shaded area in red shows *excess demand* at given prices while the green area shows *excess supply*. Compare the excesses with the curves in the supply and demand diagram in Figure 6.2. Once again, the supply curve is a pattern that shows the willingness/ability

of suppliers to put a good on the market. The demand curve shows the equivalent willingness/ability of consumers to buy the good at a given price. Putting the two together gives rise to the 'correct' price, i.e. the price at which there are no forces for the price to change. The **equilibrium price**, (or *market clearing* price) of MXN100 per DVD rental, is nothing other than the price at which there is no excess in either quantity demanded or quantity supplied – all goods that are put on the market are bought.
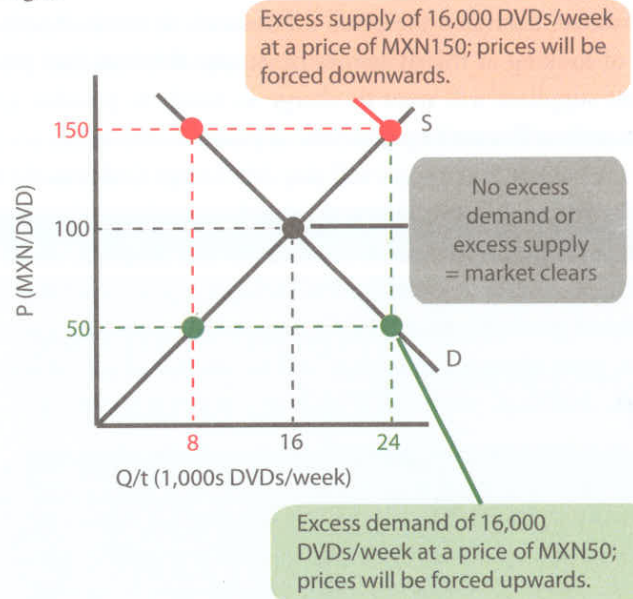

Figure 6.2: Excess supply/demand and market clearing for DVDs

Figure 6.2 shows how at *any price above MXN100* suppliers have an *incentive* to put too much on the market. The high price acts as a **signal** to suppliers to put more of the good on the market. However, consumer demand at prices above MXN100 is less than the amount suppliers would put on the market, there is *excess supply* (or surplus). Firms supplying DVDs will soon discover that they are stocking too many DVD films and will act to cut this amount. This market glut will therefore cause a downward movement in price as suppliers begin to lower prices and cut excess films on their shelves. (It is worth revising the concept of opportunity cost here. Just imagine how costly it is for a supplier of DVD films to have purchased 50 copies of Paris Hilton's latest epic drama only to discover that 49 of them sit on the shelf week after week.[1] The shelf space could be used for other items – other DVD films or complement goods such as potato chips – and the supplier's cost of obtaining the films means that these resources could have been put to better use.)

In the same manner, if the *price is set belowMXN100*, consumers will demand more DVDs than will be supplied during the

1   I wonder if the stretch of prison time for poor Paris increased demand for her DVDs. Marketing gimmick?

week. This *excess demand* will mean that suppliers have empty shelves and full queues of customers. The price is evidently too low so they will raise the price. This will act as a stimulus and incentive for suppliers to re-stock the shelves quicker in answer to the excess demand. More DVD films will be made available. An increasing market and the resulting increase in quantity supplied means that the excess demand will decrease as price moves towards *equilibrium* at MXN100 per DVD.

The above illustrates the price mechanism at work. Another way of looking at the dynamics of supply, demand and price is that suppliers will want to charge as much as possible and consumers will want to pay as little as possible. It is in this arena that the bargaining process will play out. I often find it useful to think of the supply and demand interplay as a form of auction. There are a given number of goods at an offered price and it is now up to consumers' demand to arrive at a price whereby the proffered goods are sold. Highly demanded goods will be 'bid up' in price while unsold goods will be 'auctioned off' at lower prices.

Story time!

I simply have to tell the story of my old man's (= father's) experience at an open market in Mexico City in the mid-1960s. An old lady had a large blanket spread out on the zócalo (= main plaza or square) with a largish pile of limes on it. My father, of course, needed limes for the evening cocktail party with the expatriate crowd. He asked "how much" and got a price quote per kilo and then asked "How much for the entire pile?" expecting a better deal. The old lady doubled the price! My utterly confused old man stammered out a question as to why the price was higher since he was prepared to buy them all. The answer; "Ah, señor. If I sell you all my limes I will have nothing to do for the rest of the day and I will go home. Then I won't be able to see all my friends here at the zócalo." I love the answer, for two reasons. One, it shows the limits of "market clearing theory", and two, there are people for whom there are things more important than profit. There's hope.

POP QUIZ 6.1

Market equilibrium

1. Below are some figures for the supply of wheat during a three month period. What is the equilibrium price per tonne? (Qd = quantity demanded, Qs = quantity supplied, P = price)

P ($)	Q_s (tonnes)	Q_d (tonnes)
3	2	20
4	4	16
5	6	12
6	8	8
7	10	4
8	12	0

2. What would be a possible effect on the market in question one above if the quantity demanded was higher than the quantity supplied and the price was fixed by law?

Change in Demand

The Saga (= story) continues. Recall that our so-called equilibrium price is in actual fact the price at which the market clears **within a given period of time**. What if a non-price determinant of demand were to change within the given time frame? The main non-price determinants of demand are income, price of substitutes/complements, tastes/preferences and changes in population composition.

For example, what if Swedish consumers' preferences changed in favour of more environmentally friendly modes of transportation – for example bicycles? This would shift the demand curve for bicycles from D_0 to D_1, as shown in Figure 6.3. At the new level of demand, there would be more willing buyers than supply on the market (e.g. excess demand) at P_0 and this would force prices upwards from P_0 to P_1 causing suppliers to increase the quantity supplied (Q_0 to Q_1). This is the same as saying that the increase in demand (= shift of demand curve) has caused upward pressure on the (equilibrium) price and an increase in quantity supplied. In economics we write this as "… the increase in demand (D_0 to D_1) causes an excess in demand at the original price and therefore upward pressure on the price (P_0 to P_1) and an increase in quantity supplied (Q_0 to Q_1)…"[2]

2   This is an important lesson in what economists call causality,

In economic shorthand:

- $\uparrow D_{bicycles} \to Q_D > Q_S$ at $P_0 \to \uparrow P_{bicycles} \to \uparrow Q_S \to Q_S = Q_D$ at P_1

- $\uparrow Dbicycles \to Q_D > Q_S \to \uparrow Pbicycles \to \uparrow Q_S$

- $\to Q_S = Q_D$ at P_1

This is perhaps the most important conclusion so far within our market model: *a change in demand is price-determining* – while a change in quantity demanded is price-*determined*! When demand for bicycles changes – for example, due to a change in preferences for cycling – then the entire demand curve will shift, *causing* the price to change and a *movement along* the supply curve.
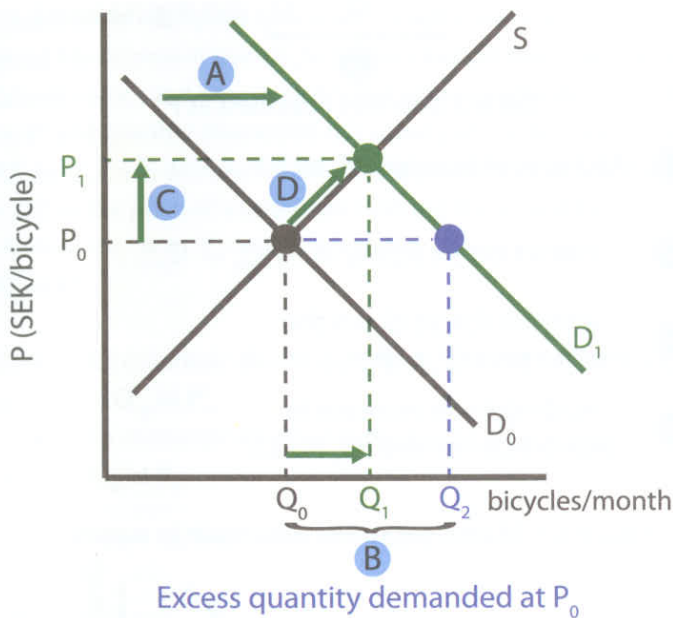
A An *increase in demand* (D_0 to D_1)…

B …causes *excess demand* at P_0 ($Q_0 \leftrightarrow Q_2$)…

C …which forces up the market *price* (P_0 to P_1)…

D …leading suppliers to increase *quantity supplied* (Q_0 to Q_1).

Figure 6.3: Increase in demand for bicycles

i.e. the forces of cause and effect. The change in demand causes a change in price – which in turn causes an increase in quantity supplied.

Naturally, there are a goodly many other non-price variables that would cause demand for bicycles to increase as in figure 6.2 but rather than spell them out for you I've inserted a…

POP QUIZ 6.2

Non-price variables and demand

1. How would demand for bicycles (in figure 6.3 above) be affected if the price of gasoline fell?

2. What other variables could affect demand as illustrated above in figure 6.3?

Story from history:

Tulipmania in 1637

The world's first proper stock market was founded in the early 17th century in Amsterdam. It didn't take too long before there was rash speculation, a commodity boom and ensuing crash – all within the time frame of less than four years.

Tulips arrived in Holland in the mid 16th century from Eastern Mediterranean countries, most likely from Turkey, where they grew in the wild. They were greatly appreciated and horticulturists set about breeding many new varieties of tulips. Many of these new types were rare and thus most coveted by the rich of Holland. They became symbols of status and sophistication in the upper class.

It didn't take long before enterprising merchants, dock workers, captains of ships, farmers and citizens from every walk of life started to speculate in an endless ascension of the value of tulips. Tulip speculation was based on the selling of bulbs (by weight). They came to be sold, speculated and

used as collateral for loans. Also, due to being highly seasonal, much of this was while the tulip bulbs were still in the ground – and nobody had actually seen the flowers as yet! Talk about a bubble – this trade actually was referred to as 'wind trade'.

Rare bulbs fetched preposterous prices as speculative euphoria and mass psychosis reached its pinnacle in 1636–37. A bulb of the Semper Augustus type worth 1,000 guilders in 1623 was sold for 5,500 guilders in 1637 – more than USD55,000 at today's prices. Traders in tulips could earn 5,000 guilders in a month where the average income at the time was 150 guilders.
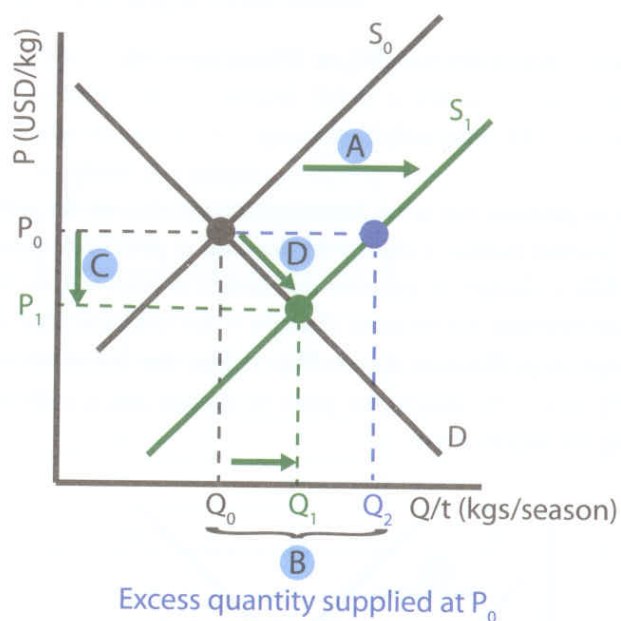
This mania came to a screeching halt during 1637 when some of the more cautious and perhaps far-sighted speculators started to leave the market by selling off their stock without replenishing it. A chain reaction set off where falling prices ultimately saw everyone trying to sell their stock before prices fell even further. The bottom fell out of the market and prices went down to under 10% of peak market values. A great many formerly wealthy people lost their fortunes to speculation as they stood after the crash with enormous debt and only tulip bulbs – now deemed worthless – as security for this debt.

Sources: M. Dash, *Tulipomania: The Story of the World's Most Coveted Flower & The Extraordinary Passions it Aroused*. Crown Publ., 1999. J.K. Galbraith, A short history of financial euphoria, 1993, Penguin Books.

- In economic shorthand: $\uparrow S_{coffee} \rightarrow Q_S > Q_D$ at $P_0 \rightarrow \downarrow P_{coffee} \rightarrow \uparrow Q_D \rightarrow Q_S = Q_D$ at P_1

Excess quantity supplied at P_0

A   An *increase in supply* (S_0 to S_1)…

B   …causes *excess supply* at P_0 ($Q_0 \leftrightarrow Q_2$)…

C   …and this forces down the market *price* (P_0 to P_1)…

D   …leading to an increase in *quantity demanded* (Q_0 to Q_1).

Figure 6.4: Market for coffee – increase in supply

In a similar vein, anyone who has experienced abnormally bad weather during the growing season for any number of vegetables and fruits can instinctively predict how the price of coffee would be forced upwards due to detrimental growing conditions and the ensuing hit to supply of green coffee. Other non-price variables which would cause the supply of coffee to increase are:

- **Scarcity/abundance of factors** of production – increased land available for coffee growing
- **Price of factors** of production – a fall in the price of farm implements or electricity
- **Technology** ('quality of factors') in production – better coffee strains which yield more crops
- **Expectations** of producers – if coffee growers expect a surge in demand in the future

Change in supply

Let us look at how a change in supply results in a change in price and thus quantity demanded. In figure 6.4 we can see how an increase in the supply of coffee, say due to a bountiful harvest of coffee beans (which is a *factor input* in the production of coffee), shifts the supply curve for coffee to the right. This is intuitively self-explanatory; an increase in market supply will create relative abundance of the good (within the time period in question) and an *excess supply* of coffee at the initial price P_0 and thus force the price downwards towards the new equilibrium (at P_1). The increase in supply (S_0 to S_1) has caused the equilibrium price to fall (P_0 to P_1), in turn leading to an increase in quantity demanded (Q_0 to Q_1).

POP QUIZ 6.3

# Effects of non-price variables on supply

1. How might the increase in supply of coffee shown in figure 6.3 above affect the market for cocoa?

2. How might cocoa producers act if they knew in advance that the supply of coffee was going to increase?

We're cooking now, right? So, how hard can it be to make a logical jump over to the outcome of a *decrease* in supply or demand? This is shown in figures 6.4a and b. In *a*, we see how supply for a good has decreased from S'_0 to S'_1 (for example due to increasing labour costs), and how the price has adjusted upwards from P'_0 to P'_1 and quantity demanded has consequently decreased from Q'_0 to Q'_1. In *b*, demand has decreased from D_0 to D_1 (say due to a fall in the price of a substitute) forcing the equilibrium price downwards from P_0 to P_1, decreasing quantity supplied from Q_0 to Q_1.

- In economic shorthand (*fig a*): $\downarrow S' \rightarrow \uparrow P' \rightarrow \downarrow Q'_D \rightarrow Q'_S = Q'_D$ at P'_1
- In economic shorthand (*fig b*): $\downarrow D \rightarrow \downarrow P \rightarrow \downarrow Q_S \rightarrow Q_S = Q_D$ at P_1

a) Decrease in supply
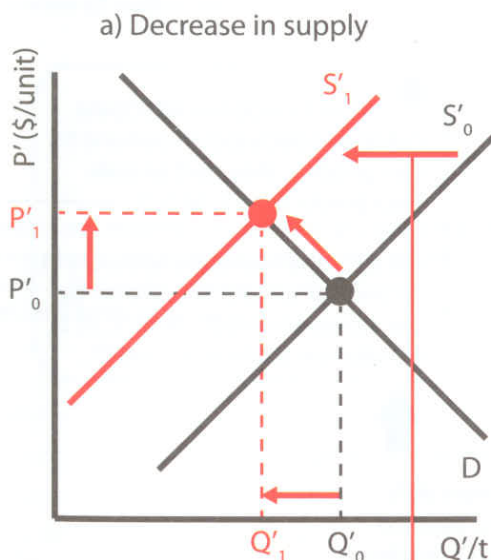
A **decrease in supply** (S'_0 to S'_1) forces the price up (P'_0 to P'_1) and thus causes a **decrease in the quantity demanded** (Q'_0 to Q'_1).

b) Decrease in demand
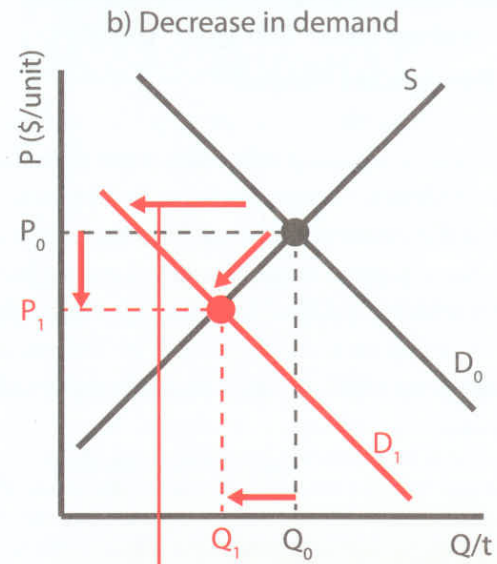
A **decrease in demand** (D_0 to D_1) forces the price down (P_0 to P_1) and thus causes a **decrease in the quantity supplied** (Q_0 to Q_1)

Figure 6.5 (a/b) : Decrease in supply and decrease in demand

DANGER MINES **WARNING !**

Increase in demand ≠ increase in supply

One of the most common errors of my fresh students is in insisting on shifting the supply curve in order to show how suppliers react to changes in demand – the explanation being that "...the increase in demand causes an increase in supply..."

The second common error is in insisting on shifting the demand curve to show how consumers react to a shift in supply, e.g. "... the increase in supply causes an increase in demand..."

Both are technically erroneous, absolutely superfluous and just plain wrong!

Any shift of the demand curve already SHOWS how suppliers will react (e.g. a change in supplied quantity) without any further assistance from the supply curve. In a similar mode, shifting the supply curve shows how consumers will react (change in quantity demanded) and no shift in the demand curve is necessary.

## Supply and demand (plus reallocation) analysis number one: 'Oh lord, won't you buy me a Mercedes Benx?'[3]

One of my IB1 kids, Nico's, has a father who is one of the largest retailers of cars in Mexico – something he no doubt regrets since I've had him put me on the short list of "Favoured Customer" for a new 425 horse power Dodge Charger.[4] I started looking for the Charger with my pal Toni last year, who basically has said "Mateo, for crying on a crutch; just get it! You can afford it…give up a new wristwatch or two." I patiently explained that

---

3   This is the title of an absolutely incredible song by Janis Joplin.
4   For reasons unknown to me, my students at some point always want to know what type of car I drive and invariably expect the answer to be either a Mercedes or a BMW. I assume it's because I wear a suit and tie but they'll probably give 'machismo' as their answer. There might be something in it. In sensible Sweden I drove very sensible cars; VW Polos and Toyota Corollas. Here in Mexico I went for speed, power and mass – an Audi A6 Turbo with 250 HP. Anything to intimidate, escape from and/or turn into purée the robo-cops (play on words; 'robo' is 'robbery' in Spanish) and wild dogs – both of which clutter up the highways when I'm enjoying complement goods. Life expectancy; T plus 2 years and counting.

I was waiting for prices to fall like a paralysed parakeet; "Toni, by the time I'm ready to buy, Nico's dad is going to *pay me* to get it off the parking lot and have two of the cute salesgirls go with me!" Am I as crazy as my colleagues say I am? Well, not in economics at least – let me explain.

The high cost of gasoline (during what will probably become known as the "Speculative oil crisis of 2008") is changing consumers' behaviour. Gasoline and cars are very close complements. As the price of gasoline surged during 2007 and 2008, households increasingly switched from large to small "Econo-cars", i.e. demand decreased considerably for large gas-guzzling cars, shown in Figure 6.6a (D_0 to D_1). This was
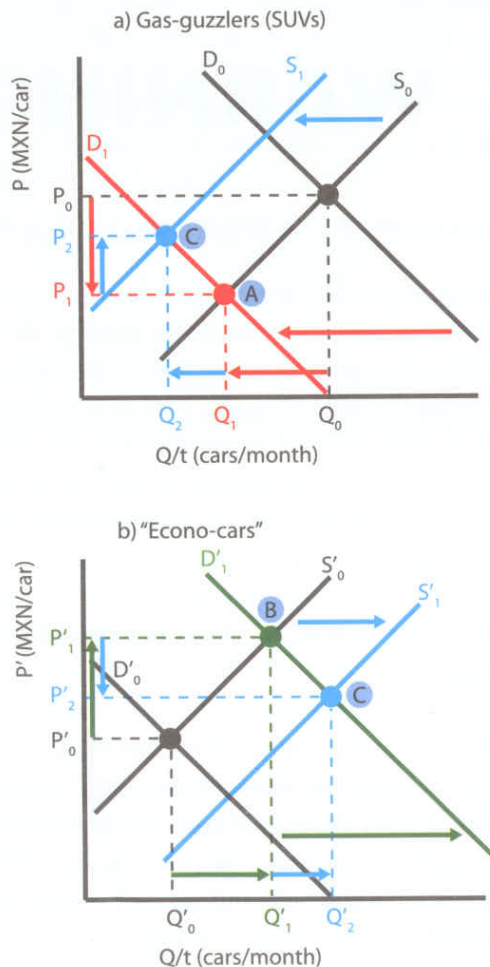
A Fig. a) As the price of gasoline rises, the demand for large fuel inefficient cars decreases (D_0 to D_1) leading to a fall in price (P_0 to P_1) and a decrease in quantity demanded (Q_0 to Q_1).

C Fig. a) As suppliers anticipate continued decreases in demand for Gas-guzzlers, they start to scale back production (S_0 to S_1)…

Fig. b) …and reallocate resources to "Econo-car" production, shown by the increase in supply from S'_0 to S'_1. Equilibrium settles at P'_2 and Q'_2.

B Fig. b) As consumers shift towards "Econo-cars", demand increases from D'_0 to D'_1 leading to upward pressure in price (P'_0 to P'_1) and an increase in quantity supplied (Q'_0 to Q'_1).

Figure 6.6: Car market – reallocation of resources

associated with an increase in demand for available substitutes, in this case more fuel-efficient cars. This is shown by the increase in demand from D'_0 to D'_1 for fuel-efficient cars in Figure 6.6b.

Now, I know what you're thinking. Didn't I – in both Figure 6.6 a) & b) – just break the rule of '…not shifting supply due to a shift in demand…'?! No – because there's more to the above story. You see, automobile manufacturers were preparing for the surge in demand by taking factors of production no longer needed in Gas-guzzlers and putting them to work in the Econo-car production plants. Resources were being *reallocated* to the production of fuel efficient cars in order to increase output during the on-going 'switch-over period'. This is shown in figure b) by the decrease in supply of gas-guzzling SUVs (S_0 to S_1) resulting in a further decrease in the quantity of SUVs purchased on the market (Q_2) and an equilibrium price of P_2. In figure b), the increase in supply from S'_0 to S'_1 for Econo-cars was the result of resources being reallocated to small fuel efficient cars. I have shown – rather speculatively, it should be noted – how the equilibrium price for Gas-guzzlers goes down due to the decrease in demand for them outstripping (= exceeding) the decrease in supply.[5] Equally speculatively, the price of Econo-cars increases since the increase in demand outstrips the increase in supply.

Basically what oil and gasoline prices are doing are reallocating resources via the laws of supply and demand. There is an enormous incentive for car makers to *divert resources* from high fuel-consuming cars to more efficient cars. This is an example of the process of reallocation.

---

5   Several of my students have gleefully pointed out that the Mexican police are in the process of changing all their old cars for new Chargers – which in fact might *increase* the price of this particular model. Our ~~bribe~~…em, *tax* money in action.

Internal Assessment tip

At this stage of your studies in economics you should be able to use basic concepts and diagrams in your Internal Assessment (IA) commentaries. I urge you to spend some time searching for appropriate articles and not simply cut something out the day before your deadline! Getting hold of a good article is half the battle.

1.   Avoid articles where the analysis is already done for you. You get no points for simple re-iteration of someone else's analysis.

2.   Don't pick overly complicated issues – there's nothing wrong with applying your acumen (= insight) and economic knowledge to day-to-day issues, such as why the municipality wishes to make public swimming pools available by lowering price of admission.

3.   It is generally more fruitful to choose a relatively brief article (not too brief!) and build up a structure of economic arguments, rather than to choose a long article and comment generally on a few definitional points. Do NOT use the small 'notices' that newspapers often put in the margins. Use proper articles.

4.   Diagrams are to be used to support the iteration – not replace it! Make sure that your diagrams are neat, numbered and well referred to in the text and not simply left hanging. This makes it easier for your teacher to follow your thread of thought.

Don't waste time going through stacks of newspapers/journals but be smart. Do a "Google" and simply type in a search string containing key words which might be dealt with via your knowledge of economics at this stage. For example, during 2007/'08 I suggested my people try "…gasoline + oil + demand + supply + tortillas + maize + solar power + wind power + silicon…". Wonder what I'm on about?! Read the next section on "Tortillas and oil".

Supply and demand analysis number two. 'Tortillas and oil'

"While many worry about filling their gas tanks, many others around the world are struggling to fill their stomachs. And it's getting more and more difficult every day."

World Bank President, Robert Zoellick, (The Guardian, 30 April 2008)

It was applied economics in the making; we watched, week by week, how oil prices wreaked havoc (= chaos) on poor families in Mexico. No, not because of rising gasoline prices – gasoline is heavily subsidised in Mexico – but because of ever higher prices for maize and maize tortillas. World food prices – according to the International Monetary Fund (IMF) – rose by some 45% between the end of 2006 and April 2008. In Mexico, for the 50% of the country's 106 million citizens living on less than five US dollars a day, the staple (= basic, essential) food is the tortilla – a flat "pancake" like bread made of maize meal (corn flour). When the price of maize skyrocketed during 2007/'08 the effect was to cause the price of tortillas to increase from around 7 Mexican pesos (MXN) to over MXN10.[6] I didn't notice the price increase – but I don't live on USD5 a day as millions of Mexicans do.[7] I did, however, notice the thousands of protesters in Mexico City banging empty pots against the walls of government buildings.

So what caused the rise in the price of tortillas? A number of things – all related. The main cause was the surge in oil and gasoline prices which caused a somewhat frantic search for substitutes. One of these substitutes is ethanol (alcohol) made from, primarily, sugar and maize. US farmers started shifting agricultural land (standard example of reallocation of resources!) towards the production of maize and increasingly sold this for use in ethanol production. Hence, as maize meal for tortillas and maize for ethanol are in joint supply, an increase in demand for ethanol in the US served to decrease the available supply of maize for foodstuffs. How could this affect Mexico which supplies most of the maize for its domestic market? Three causes stand out: *competition laws*, higher *import prices* and *speculation* – but not necessarily in that order of occurrence or importance. Let me do a brief point-by-point explanation:

---

6 The exchange rate then was between MXN10.5 to MXN11 for 1USD.

7 The official minimum wage in Mexico (at the current exchange rate of USD1 = MXN13.5) is just over USD4. It rather makes sense why my Mexican ex-lady friend refused to get in the car with me if I was wearing one of my watches. Basically, my Rolex GMT Master II represents about 6 years minimum wage and she didn't want to be a collateral victim of highway robbers due to my nonchalant habit of letting my left arm hang out the car window during traffic. Affectionate parents who invite me to parties inevitably send a car and driver for me since I refuse to wear a cheapo watch. Maybe they figure I need both hands to do term reports.

1. Mexico has allowed imports of maize since 1994[8] but has *limited* these imports to protect domestic farmers from the very efficient and heavily subsidised American farms. However, 80% of imported maize in Mexico comes from the US. As fuel prices in the US soared and ethanol demand subsequently increased (recall that an increase in the price of a good will increase demand for substitutes) more and more US maize was diverted towards ethanol, creating a surge in maize prices.

2. This decreased the US supply of maize for exports and increased the price of imported maize to Mexico.

3. The main maize suppliers for tortillas in Mexico are, somewhat notoriously, operating in an oligopoly market. There is ample evidence that domestic Mexican suppliers withheld maize in speculative anticipation of higher future prices.

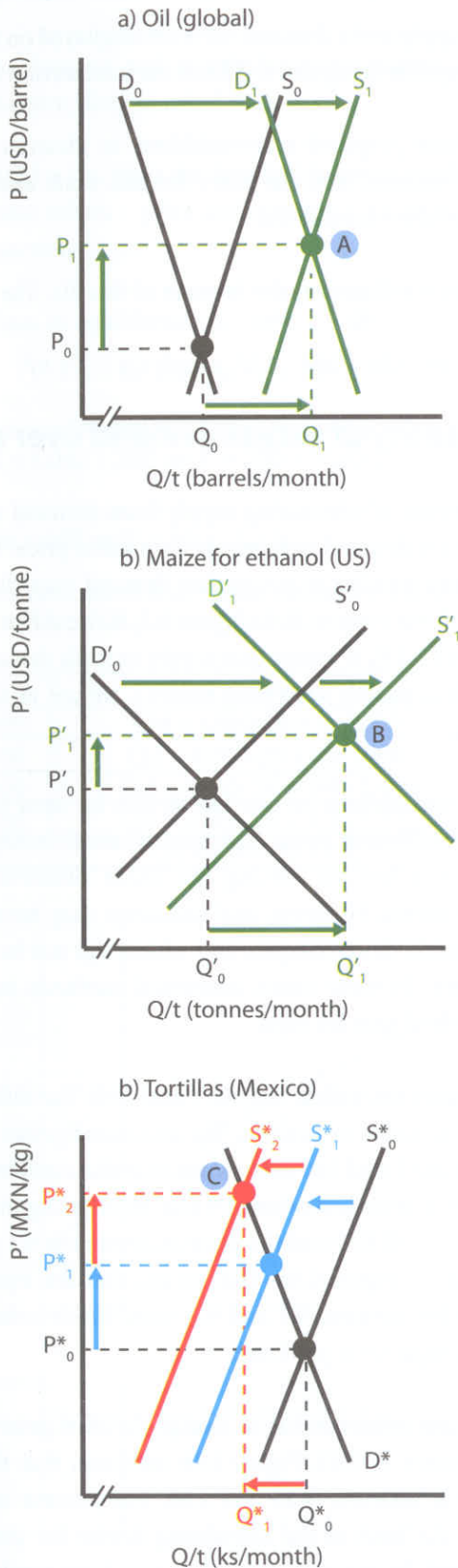The causal flows (oil → maize → tortillas) are illustrated in Figure 6.7

- Demand for oil increased (figure a) during 2006 – 2008 due to increasing demand from oil hungry countries – primarily India and China – and speculation of higher prices. While suppliers attempted to increase production of oil, demand far outstripped supply. This caused gasoline prices to double in the US during the time period and consequently caused a search for substitutes – primarily ethanol.

- This in turn caused an increase in demand for maize to produce ethanol. It also caused maize growers to step up maize production to meet ethanol demand. This is illustrated in figure b by the increase in demand from D'_0 to D'_1 and the increase in supply from S'_0 to S'_1. Since more US maize was now going to the production of ethanol, the available supply of maize for the Mexican market decreased and Mexican import prices for maize shot up.

- This resulted in higher production costs for Mexican tortilla producers, shown by decrease in supply (S^*_0 to S^*_1) in fig c. When maize producers in Mexico acted upon *expectations* of higher world market prices for maize, they started to hoard supplies, e.g. they held off on releasing enough maize to meet current demand in speculation of higher future prices.

---

8 When the North American Free Trade Agreement came into effect.

The decrease in maize flower (factor input in tortillas) further increased production costs for tortilla producers, decreasing supply (S^*_1 to S^*_2) and drove up the price of tortillas by some 200% during the first months of 2007.[9]

9   See http://www.nytimes.com/2007/01/19/world/americas/19tortillas.html, http://www.reuters.com/article/GCA-Agflation/idUSN1451427920080514 and http://news.bbc.co.uk/2/hi/americas/7432164.stm

a) Oil (global)

A   As global demand for oil drastically outstripped supply during 2006 – 2008, the price of oil and thus gasoline in the US rose.

b) Maize for ethanol (US)

B   The search for available substitutes – here, ethanol – caused an increase in demand for maize (D'_0 to D'_1), in excess of the increased supply (S'_0 to S'_1), causing an increase in the price of maize (P'_0 to P'_1).

b) Tortillas (Mexico)

C   As more US maize was diverted towards ethanol production, available imports to Mexico of maize from the US decreased which in turn increased costs for tortilla producers. The supply of tortillas decreased (S^*_0 to S^*_1) …

…and at the same time, speculative forces in Mexico lead to hording of maize amongst maize suppliers raising maize prices further, and a further decrease in the supply of tortillas (S^*_1 to S^*_2.) Tortilla prices had trebled by January 2007 (P^*_2).

Figure 6.7: Oil, maize and tortillas in Mexico, 2006 – 2008

Here it was my intention to continue with the secondary and tertiary effects on the markets for solar power, silicon and computers. But I'm tired, so I shall instead turn to the trusted method of didactic (= educational) laziness etched in stone tablets by countless generations of teachers since the days of Ivied Castles all around, hi-ho. Yes, you guessed it; I will pose more questions.

POP QUIZ 6.4

Supply and demand

1. Re *Tortillas and oil* above: How might the price of oil have affected the price of solar panels in the US?

2. Re *Tortillas and oil* above: How might the price of oil have affected the price of silicon? (This is a continuation of Question 1 above. Hint; look up 'photovoltaic cells' and 'silicates'.)

3. Re *Tortillas and oil* above: How might the price of oil have affected the supply of computers? (This is a continuation of Question 2 above. Did you look things up?)

4. 'In some countries the price of houses rises while the demand for houses rises'. Using supply and demand analysis, explain why this statement does not contradict the law of demand.

5. Using supply and demand diagrams, explain how the price of seasonal goods would be prone to fluctuate during a year. ´

6. You believe that property prices will soon plummet (= fall drastically). Being a property-speculator, you act on this belief. Illustrate the possible outcome of this situation using supply and demand curves. Assume that other speculators will follow your lead.

Plotting demand and supply curves

As discussed earlier, individual demand (and/or supply) curves can be drawn using tables of values and graph paper, spreadsheets or graphic calculators.

If a supply and a demand curve are displayed on the same graph, the equilibrium point will be at the intersection of the two lines.

We have prepared a spreadshheet to illustrate this. Visit the IBID website, find the entry for this book and download the spreadsheet Equilib.xls.

Figure 6.8 illustrates the features of this file. The starting curves are:

Demand: $Qd = 200 - 12P$, supply $Qs = 7 + 9P$

Quantity of excess demand over supply

The result of subtracting supply from demand is known as the excess of demand over supply for a given price. If this result is a positive number, it means that demand exceeds supply. This is the case in rows 4-11 in Figure 6.7. If the difference is negative (rows 12-23), it means that supply exceeds demand. Where this result is exactly zero (not shown), we are at the equilibrium point.

This spreadsheet, or one like it, can be used to 'home-in' on the equilibrium point. The mathematical technique described earlier is best for solving the linear situations described in this chapter. However, real situations may have curves which are much more complex and which will not be solvable using algebra. In these cases, numerical methods such as the one described here are used.

In Figure 6.8, look at cells G11 and G12. The difference changes from negative to positive. The associated prices (cells D11 and D12) are 9 and 10. This means that the equilibrium point lies between price lies between 9 and 10. Altering the price starting point in cell B12 and the price increment to 0.1 gives the result shown in Figure 6.8a. This means that the equilibrium point price lies between 9.1 and 9.2 (note where column G changes from negative to positive).

The next refinement is to change the start price to 9.1 and the increment to 0.01. Figure 6.9b indicates that the equilibrium price is between 9.19 and 2.00. The process can be repeated until the limit of the calculáting device (or your patience) is reached. In practice, stop when you have reached the desired level of accuracy.

HL extensions

Equilibrium price and quantity – linear supply and demand functions and calculating equilibrium price and quantity

P	Qs ($)	Calculation
0	-2,000	Qs = -2,000 + 40 × 0; -2000
50	0	Qs = -2,000 + 40 × 50; 0
100	2,000	Qs = -2,000 + 40 × 100; 2,000
150	4,000	Qs = -2,000 + 40 × 150; 4,000
200	6,000	Qs = -2,000 + 40 × 200; 6,000

Now we put the demand and supply functions together and calculate equilibrium. Keep in mind that:

a)   If Qd = Qs, then the demand function yields the same values – price and quantity – as the supply function

b)   Thus, in equilibrium; a – bP = c + dP

$Qd = a – bP$ and $Qs = c + dP$

Assume the following demand and supply functions:
$Qd = 4,000 – 20P$ or $P = 200 – \frac{1}{20}Qd$

and $Qs = -2,000 + 40P$ or $P = 50 + \frac{1}{40}Qs$

Putting this into a table shows that market equilibrium is established at a price of $100 and quantity of 2,000 units.

P	Qd ($)	Calculation
0	4,000	Qd = 4,000-20 × 0; 4,000
50	3,000	Qd = 4,000-20 × 50; 3,000
100	**2,000**	**Qd = 4,000-20 × 100; 2,000**
150	1,000	Qd = 4,000-20 × 150; 1,000
200	0	Qd = 4,000-20 × 200; 0

Basically, finding market equilibrium is a matter of solving one unknown, for example P. Thus solving the following equation ('simultaneous equations' in math speak) will find P:

- Assuming market equilibrium, where Qd = Qs, we equate the equations to get **4000 - 20P = -2000 + 40P**[1]

  o $6000 = 60P; P = \frac{6000}{60}; P = 100$

- Now we solve for Qd or Qs by simply substituting P = 100 in either function:

  o Qd = 4000 – 20 × 100; 2,000 units

  o Qs = -2000 + 40 × 100; 2,000 units

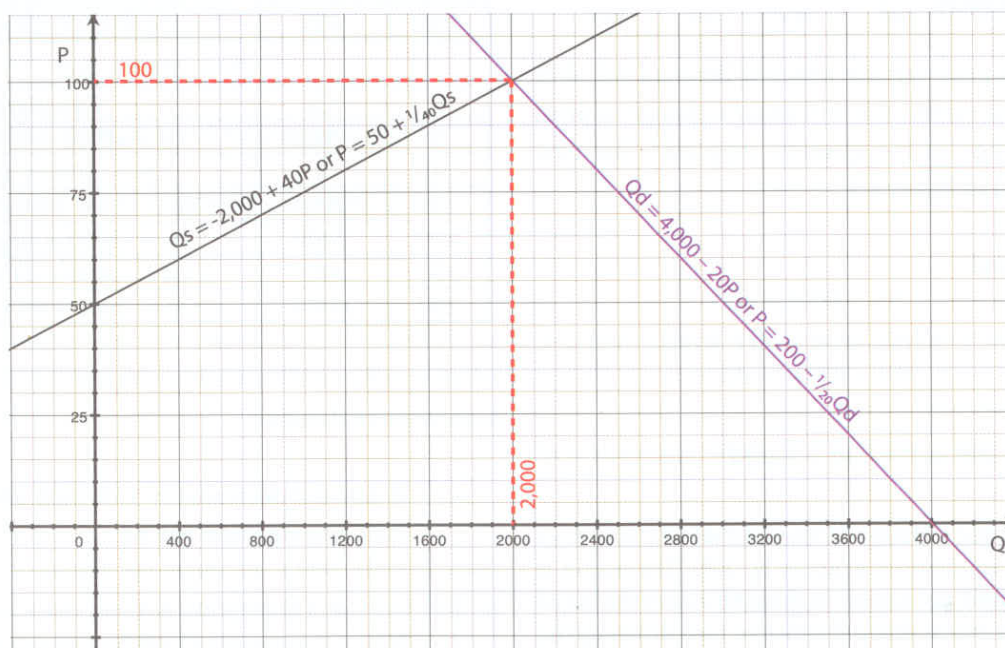The supply and demand curves are plotted out in Figure 6.8.

_____

1   For the highly mathematically inclined, you can take the simultaneous equation a – bP = c + dP and collect terms.

Thus $a – c = P(d + b)$ and $P = \frac{(a – c)}{(d + b)}$.

Figure 6.8 Market equilibrium

Plotting demand and supply curves; excess demand/supply

At any other price than $100, there will be an *excess supply or demand* on the market – shown in Figure 6.9, where $150 and $75 are non-market clearing prices.

Calculating excess supply (Qd = 4000 – 20P and Qs = -2000 + 40P)

- At P(150) we would get

    o Qd; 4000 – 20 × 150 = 1000

    o Qs; -2000 + 40 × 150 = 4000

The excess supply is 3000 units at P(150)

Calculating excess demand (Qd = 4000 – 20P and Qs = -2000 + 40P)

- At P(75) we would get

    o Qd; 4000 – 20 × 75 = 2500

    o Qs; -2000 + 40 × 75 = 1000

The excess supply is 1500 units at P(75)

Plotting demand and supply curves; a change in demand

Any change in a non-price variable affecting demand will shift the curve. As outlined earlier, this will cause a change in price and therefore a change in the quantity supplied.[2] The new equilibrium is plotted in Figure 6.9.

An increase in demand

If demand increases by 15% (600 units at all P levels), then new equilibrium will be:

- **Qd = 4600 - 20P or P = 230 – $\frac{1}{20}$Qd**

---

2 It is worth noting that a change in demand of, say, 20% does NOT change final equilibrium quantity by 20%. The reason is rather straightforward; an increase in demand causes the initial price to result in an excess demand which in turn drives up the price. As the price rises, the Qs increases and the Qd decreases – final equilibrium will depend on the elasticities of supply and demand.

- Qs = -2000 + 40P or P = 50 + $\frac{1}{40}$Qs

- **New equil/mkt clearing: 4600 - 20P = -2000 + 40P**

    o **Solving for P**: 6600 = 60P, P = 110

    o **New Qs**: -2000 + 40 × 110 = 2,400

    o **New Qd**: 4600 – 20 × 110 = 2,400

- Again, to get the P-intercepts:

    o For the D-curve you simply divide 'a' by 'b', e.g. 4,600÷20 = 230

    o For the S-curve (which *will* come in handy when we calculate tax incidences and such!), same thing; 'c' divided by 'd', so 2000÷40 = 50 (Note that it is entirely possible to get a minus value for the P-intercept since the S-curve might intercept the Q-axis first! If 'c' is negative then the P-intercept will be a positive value. If 'c' is positive, the P-intercept will be negative.)

Increase in demand of 15% (e.g. 600 units at ALL prices)
Qd = 4600 - 20P

P	Qd	Calculation
0	4600	Qd = 4,600-20 × 0; **4,600**
50	3600	Qd = 4,600-20 × 50; **4,000**
100	2600	Qd = 4,600-20 × 100; **3,600**
110	2400	Qd = 4,600-20 × 110; **2,400**
150	1600	Qd = 4,600-20 × 150; **1,600**
200	600	Qd = 4,600-20 × 200; **600**

Qs = -2000 + 40P

P	Qs	Calculation
0	-2000	Qs = -2000 + 40 × 0; **-2000**
50	0	Qs = -2000 + 40 × 50; **0**
100	2000	Qs = -2000 + 40 × 100; **2,000**
110	2400	Qs = -2000 + 40 × 50; **2,400**
150	4000	Qs = -2000 + 40 × 75; **4,000**
200	6000	Qs = -2000 + 40 × 100; **6,000**

A decrease in demand

- If demand decreases by 15% (600 units at all P levels), then new equation will be:

    o **Qd = 3400 - 20P**

    o **Qs = -2000 + 40P**

New equation: 3400 - 20P = -2000 + 40P

Solving for P: 5400 = 60P, P = 90

- ○ **New Qs:** -2,000 + 40 × 90 = 1600

- ○ **New Qd:** 3,400 – 20 × 90 = 1600

The intercepts:

- P-intercept for the D-curve is $\frac{3400}{20} = 170$

- P-intercept for the S-curve is $\frac{2000}{40} = 50$

Decrease in demand of 15% (e.g. decrease in Qd of 600 units at ALL prices)

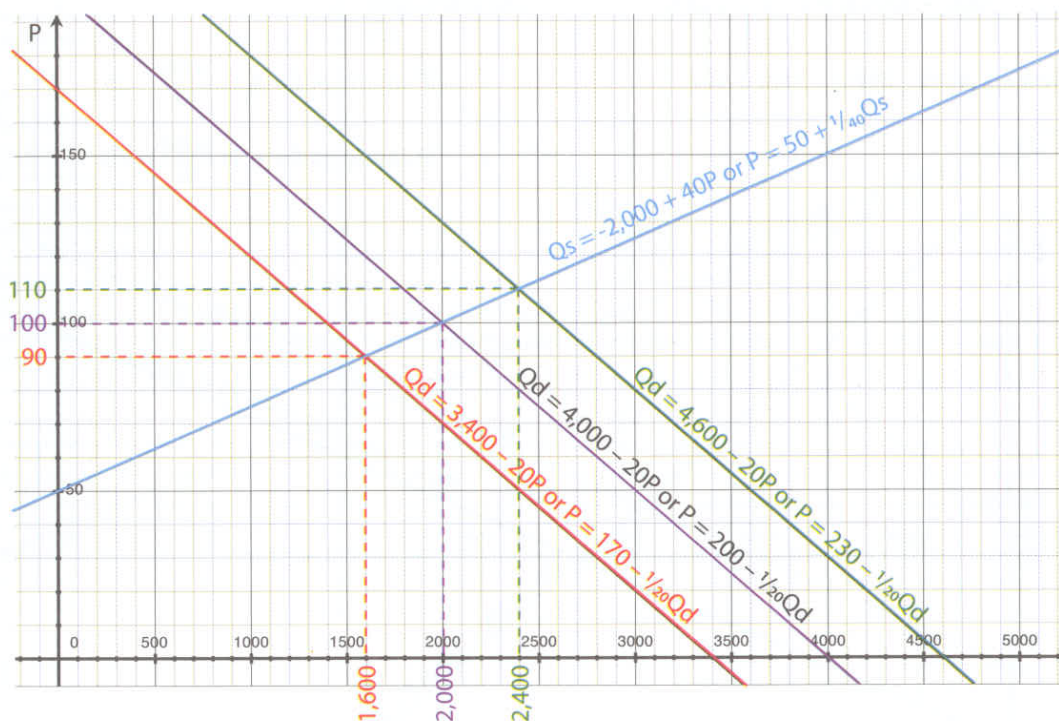| Qd = 3400 - 20 × P		Qs = -2000 + 40P	
P	Qd	P	Qs
0	3400	0	-2000
50	2400	50	0
90	1600	90	1600
100	1400	100	2000
150	400	150	4000
170	0	200	6000

Figure 6.9 Shift in demand and new equilibrium

Plotting demand and supply curves; a change in supply

A change in a non-price determinant of supply will shift the supply curve, causing a change in price and thus a change in the quantity demanded.

A change in supply

- **Increase**; if supply increases by 1,000 units at each price level, then there is an upward (parallel) shift of the supply curve and the Q-intercept is -1,000.

- ○ **Qd = 4000 – 20P**

- ○ **Qs = -1000 + 40P**

New equation: 4000 - 20P = -1000 + 40P

- **Solving for P:** 5000 = 60P, P = 83.3

- ○ **New Qs:** -1,000 + 40 × 83.3 = 2333

- ○ **New Qd:** 4000 – 20 × 83.3 = 2333

Qd = 4000 - 20 × P		Qs = -1000 + 40 × P	
P	Qd	P	Qs
0	4000	0	-1000
50	3000	50	1000
83.33	2333.4	83.33	2333.2
100	2000	100	3000
150	1000	150	5000
200	0	200	7000

- **Decrease**; if supply decreases by 1,000 units at each price level, then there is a downward (parallel) shift of the supply curve and the Q-intercept is -3,000.

  o $Qd = 4000 - 20P$ or $P = 200 - \frac{1}{20}Qd$

  o $Qs = -3000 + 40P$ or $P = 75 + \frac{1}{40}Qs$

  o New eq: $4000 - 20P = -3000 + 40P$

- Solving for P: $7000 = 60P$, $P = 116.67$

  o New Qs: $-3000 + 40 \times 116.67 = 1666.7$

  o New Qd: $4000 - 20 \times 116.67 = 1666.7$

Qd = 4000 – 20 × P		New supply curve: Qs = -3000 + 40P	
P	Qd	P	Qs
0	4000	0	-3000
50	3000	50	-1000
100	2000	100	1000
116.67	1666.6	116.67	1666.8
150	1000	150	3000
200	0	200	5000

The intercepts:

- P-intercept for S_1 is $\frac{1000}{40} = 25$

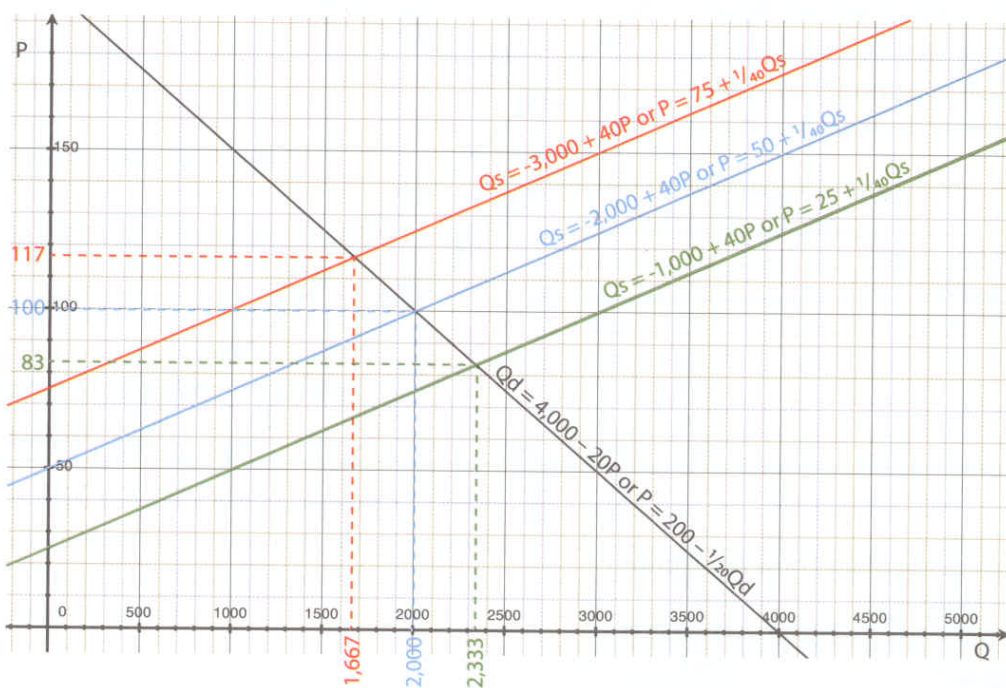- P-intercept for S_2 is $\frac{3000}{40} = 75$

Figure 6.10

Slope: a change in d in our S-function of $Qs = c + dP$ will of course result in a change in the slope

- Assume that the new supply function is $Qs = -2000 + 50P$ or $P = 40\frac{1}{50}Qs$

- Basically, this tells us that for any change in P of $1, the Qs will increase by 50

  o An increase in price of 25 → $\Delta\uparrow Qs$ of 1,250

  o An increase in price of 50 → $\Delta\uparrow Qs$ of 2,500

  o The Q-intercept is still -2000 but the P-intercept is $\frac{2000}{50}\left(\frac{c}{d}\right) = 40$

New equation: $4000 - 20P = -2000 + 50P$

- **Solving for P**: $6000 = 70P$, $P = 85.71$

  o New Qs: $-2{,}000 + 50 \times 85.71 = 2285.5$

  o New Qd: $4000 - 20 \times 85.71 = 2285.5$

Supply decrease in slope:

$Qd = 4000 - 20 \times P$

$Qs = -2000 + 50P$

P	Qd	P	Qs
0	4000	0	-2000
50	3000	25	-750
85.71	2285.8	85.71	2285.5
100	2000	50	500
150	1000	75	1750
200	0	100	3000

Figure 6.11

Paper 3 example for HL (25 marks in total)

1. The weekly supply and demand curves for train fares between Hamburg and Lübeck are given by

    $Qs = -40 + 30P$

    $Qd = 200 - 10P$

    This is the quantity supplied and demanded for train tickets per week in Euros (€).

    a. Calculate Qs and Qd on the market of setting the price at €9 per ticket. *[2 marks]*

    b. Fill in the axes and scale in a diagram. Questions (c) to (i) link to this diagram. *[1 mark]*

    c. Draw the supply and demand curves in the graph and indicate the P intercept for the supply curve and quantity intercept for the demand curve. *[3 marks]*

    d. Calculate equilibrium price and quantity and identify market equilibrium in the diagram. *[4 marks]*

2. A Boeing 747 from the Swedish airline Störting Airways goes down in the Baltic causing widespread panic amongst European air passengers. Demand for train tickets increases by 40% per week at all price levels.

   a. State the new equation for the demand curve.

      [1 mark]

   b. Draw the new D-curve in the diagram clearly identifying the Q intercept   *[2 marks]*

   c. Explain, using correct economic terminology and references to the diagram, why the change in demand yields a new equilibrium price.

      [4 marks]

   d. Calculate the new equilibrium price and quantity.   *[2 marks]*

   e. Explain, using your calculations and the curves in your diagram, what would be the effect of a minimum price set at the original equilibrium price.   *[4 marks]*

3. Assume that in the original equilibrium, a subsidy to rail companies caused a new supply function of

   $$Q_s = -10 + 30P$$

   a. Calculate the € amount of the subsidy per unit.

      [2 marks]

Summary & revision

1. Excess demand is when the price is below market equilibrium and quantity demanded is greater than quantity supplied.

2. Excess supply is when the price is above market equilibrium and quantity supplied is greater than quantity demanded.

3. When the quantity demanded equals the quantity supplied during a given time period, the market has cleared and there is market equilibrium.

4. An increase in demand forces up the market price and leads to an increase in the quantity supplied.

5. An increase in supply forces the market price down and leads to an increase in the quantity demanded.
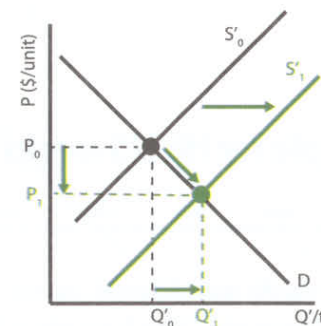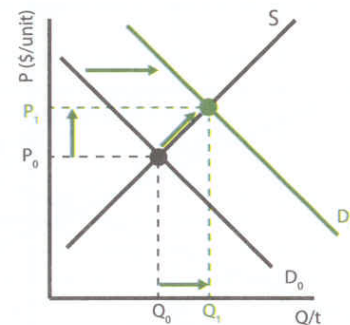
HL extensions

6. Equilibrium price (linear demand function).

7. Equilibrium quantity.

8. Plotting linear supply and demand curves.

9. Excess supply and demand.

7. The Price Mechanism and Resource Allocation

Key concepts:

* The basic economic problem – revisited
* Choice and opportunity cost
* The price mechanism and resource allocation

"An economist is an expert who will know tomorrow why the things he predicted yesterday didn't happen today." Laurence Peter

The basic economic problem – revisited

The basic economic problem (see Chapter 2) of "what to produce, how to produce it and for whom the output is produced" is a question faced by all economies. Two basic solutions to this have been put forth historically, a planned economic system such as the USSR and a market based system. Neither of these ever existed in a pure form – instead most economic systems were mixed economies with a degree of planning and markets side by side. Here we shall briefly look at the basic issue of how the factors of production – land, labour, capital and the entrepreneur – are allocated in a market system by way of the **price mechanism**.

Choice and opportunity cost

Since societal wants are unlimited (infinite) and factors of production are limited (finite), choices must be made – by both consumers and suppliers. When I buy the Rolex Deepsea for USD8,500 I am giving up…well, I am giving up whatever is my *next best alternative*. This is the trade-off in my choosing the Rolex, the foregone opportunity of a three week luxury vacation or three new tailored Hong Kong suits – whichever of the two foregone options is my preffered. Extra work writing textbooks means foregone time spent at the beach; additional hours after school helping students means less time in the gym, and so on. Every choice comes with an opportunity cost price tag attached.

When economists speak of costs in terms of supplying wants, we are in fact referring to foregone opportunities in using limited resources – the **opportunity costs** of production. We are also speaking of **economic goods** – e.g. goods which are scarce since there are alternative uses. In economics we often use a production possibility frontier to illustrate opportunity costs.
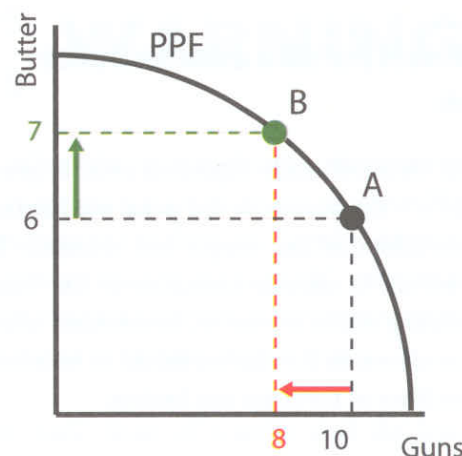

Figure 7.1: Opportunity costs and the PPF

Assuming only two goods, military goods and civilian goods, an economy (either by planning or market forces) can put all available factors into the production of one or the other – and any combinations along the curve. In Figure 7.1 we assume that the economy is operating at maximum output (e.g. the economy is operating on the PPF) so it is impossible to re-allocate resources from military goods to civilian goods without incurring the opportunity costs of fewer military goods. In the example a movement from 6 units of butter to 7 units of butter means an opportunity cost of two units of guns. All goods 'compete' for scarce resources. In some way society will make a choice in the re-allocation of resources from certain goods to others.

This is a USD30 pile of wood that Rob the housey and I bought for the household in Mexico City during the winter of 2008/09. Since there are few forests left here in Mexico we have a classic example of scarcity and high price. And yes, we rationed our consumption and instead went on to steal fallen wood from the national park next door.

The price mechanism and resource allocation

So how is the 'decision' made regarding what to produce and how much of it? The answer is the *price mechanism* arising from the interaction of the supply and demand. The price mechanism serves to allocate resources to the 'best' use in terms of allocating scarce resources. Economists often refer to the dynamism shown in the market model in terms of the two signalling functions of the price mechanism:

1. The **incentives function**

2. The **rationing function**

Any change in market price is a **signal** to suppliers to either increase or decrease output and to consumers to increase or decrease consumption. This is perhaps one of the great successes in the market system – and shows the dismal failure of a centrally planned system. An increase in price means increased ability/willingness of suppliers to put more on the market – this is the **incentives** function of the producers on the market; more will be provided if demand drives up the price of the good, owing to the driving forces of firms' self-interest.

Of course, a higher price for goods is also a signal to consumers – scarcity of goods drives up the price and available goods need to be **rationed** in some way. Consumers will be discouraged

from consumption to a certain extent and decrease their demand.

A tricky question in our market model deals with the speed at which the market clears – or whether the market indeed clears at all! This is highly dependant on the market characteristics. An immediate shock to the market for any number of goods might lead to excess demand or supply for an indeterminate period. For example, it is highly unlikely that a decrease in the supply of beef would *immediately* cause an increase in the price of hamburgers – even in the most open and responsive economic systems. However, our model and subsequent analysis indicates that in the longer term, market disequilibrium cannot exist and that prices will indeed move in order to balance relative scarcities or excesses. The price tends to *move towards* market equilibrium.

Taken outside one of the lecture halls at the Oxford College where I do revision courses for the OSC. Now, do you think this is good resource allocation?

Definition: 'Price mechanism', 'signalling function', 'incentives function' and 'rationing function'

The price mechanism is the market mechanism whereby scarce resources are allocated between competing uses. The price has a signalling function – a higher price means that suppliers have an incentive to increase supply whereas consumers are signalled that they should consume less. The price signal also has an incentives function for firms – a higher price is an incentive for producers to increase supply in order to increase profit. Accordingly, the price also serves a rationing function since a higher price will discourage demand.
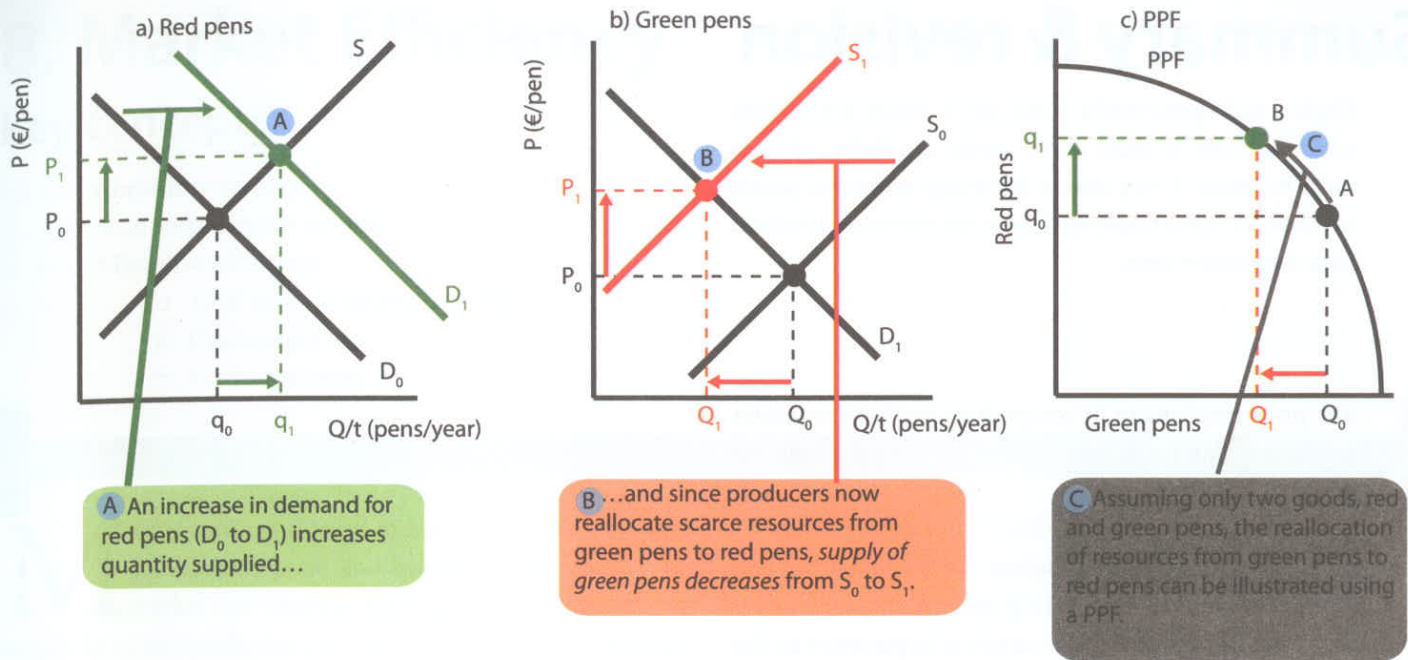
Figure 7.2 Illustrating re-allocation

Within the figure:

a) Red pens — P ($\unicode{8364}$/pen) vs Q/t (pens/year); S, D$_0$, D$_1$; points P_1, P_0, q_0, q_1; A

A An increase in demand for red pens (D_0 to D_1) increases quantity supplied…

b) Green pens — P ($\unicode{8364}$/pen) vs Q/t (pens/year); S_1, S_0, D_1; P_1, P_0, Q_1, Q_0; B

B …and since producers now reallocate scarce resources from green pens to red pens, *supply of green pens decreases* from S_0 to S_1.

c) PPF — Red pens vs Green pens; PPF; points A, B, C; q_1, q_0, Q_1, Q_0

C Assuming only two goods, red and green pens, the reallocation of resources from green pens to red pens can be illustrated using a PPF.

Figure 7.2 a) to c) illustrates how the flexibility of competitive markets leads to market dynamism where prices change – ultimately – in accordance with the forces of supply and demand. Assume two goods, for example red and green pens, and optimum efficiency (*Pareto optimum* – see next chapter) in an economy. Recall the basic issue of scarcity and endless wants and how markets solve the basic economic problem via the **signalling function of prices**.

- Point A in **Figure a)** illustrates an increase in demand for red pens which drives up the price and signals suppliers (*incentive function of price*) to increase the quantity supplied of red pens (q_0 to q_1).
- Point B, **Figure b) and c):** Since the economy is operating at optimum efficiency (point **A**, **Figure c)**), suppliers must reallocate resources from green pens to red pens, points **A** to **B** in the PPF, **Figure c)**.
- As resources used in the manufacturing of green pens become more scarce, the supply for green pens decreases from S_0 to S_1 (**Figure b)**). The increase in the price of green pens (P_0 to P_1) signals consumers to decrease their consumption (*rationing function of price*) of green pens – quantity demanded for green pens decreases from Q_0 to Q_1.

DANGER MINES # WARNING !

The issue of how the price mechanism allocates resources is one of the more common exam questions posed in IB economics. It is also somehow one of the most sloppily addressed. Variations of this question are along the lines of "…explain how the price mechanism allocates resources in a market economy…" and ask that you address three fundamental issues:

1. The basic issue of scarcity and the basic economic problem

2. How the price acts as a mechanism to allocate resources

3. Use supply and demand analysis to illustrate how prices act as signals to producers (incentives function) and consumers (rationing function).

I strongly urge you to get a good, solid grip on the diagram series in Figure 7.2.

Summary & revision

1. Choice and opportunity costs: since wants are infinite and resources to fulfil these wants are finite, choices must be made. Every choice involving economic goods involves an opportunity cost since the resources involved have alternative uses.

2. The price mechanism in competitive market economies serves to allocate scarce resources towards different goods.

   a. The signalling function tells suppliers and consumers whether to produce/consume more – a higher price signals suppliers to put more on the market but signals consumers to consume less.

   b. The incentives function of price means that suppliers will increase output at higher prices since there is a profit incentive to do so.

   c. The rationing function of price serves to discourage consumption of goods when the price increases.

8. Market Efficiency

Key concepts:

- Consumer surplus
- Producer/supplier surplus
- Allocative efficiency
  - Total surplus: societal surplus
  - Deadweight loss
  - *Pareto optimum*

Markets are considered to have "done the job" when equilibrium price and quantity is achieved – the market has cleared, leaving on excess demand or supply. In other words, the price mechanism has seen to it that resources have been allocated to the right areas and consumers are paying the right price for the correct amount of goods.[1]

Consumer surplus

I always enjoy this part of the syllabus in class as it lends itself very easy to students' own experiences and intuitiveness. I ask you too to imagine walking by a store window on the way to school every day, and seeing a really nice jacket that you just must have. It's expensive, €180, so you save for a few months. One day you walk in with your wad of money to buy the jacket, and it turns out that you have just hit the sales! The jacket 'only' costs €150. You buy it and enjoy the additional pleasure of having been prepared to pay €180 but not having to do so. You get a 'surplus' valued at €30, i.e. an enjoyment above and beyond what you paid for. You got €180 worth of marginal utility for €150.[2] This is **consumer surplus**.

Let's assume a local market for jackets and that each potential customer only wants one jacket and that demand is zero at a price of €210. Figure 8.1 illustrates how consumer surplus is summed up.

- **Areas A to E**: Let's say that 100 people walk in to buy a jacket and are prepared to pay a price of €200 and the market price is €150. This "group" gets a marginal utility of areas A, B, C, D and E; 100 units time €50.

- **Areas F to I**: If 200 people buy a jacket each, the first group of 100 people (Group A to E) will receive a surplus benefit (i.e. pay a price below the value they place on the item) of €50 each. The second group, F to I, is prepared to pay €190 yet also pay the market price of €150. The consumers in this group thus receive a surplus of €40 each.

At the market price of €150, total consumer surplus is shown by the fifteen shaded squares. And as prices in fact do not move in increments of €10, but can theoretically move 'one cent at a time', total consumer surplus in the entire orange shaded triangle above the market price, shown in Figure 8.1. Conclusion: At a market price of €150, a possible 599 people are getting 'more than they pay for', i.e. receive consumer satisfaction valued at more than the market price. The concept of consumer surplus tells us that all economic transactions are of value to participants on the consumption side – otherwise they would not take place. Any intervention (= non-market force) on a market which diminishes consumer surplus has a negative effect on both consumers and allocative efficiency.

> **Definition: 'Consumer surplus'**
>
> **Consumer surplus** is the additional benefit (marginal utility) received by a consumer in purchasing a good where the market price is less than he/she would be willing to pay. Total consumer surplus is the area above the market price but below the demand curve.

---

1 A more correct way of putting it in economic jargon is that the marginal benefit to consumers (quantity demanded at equilibrium) is equal to the marginal cost (quantity supplied at equilibrium) of producing the goods. This will be dealt with in Chapter 16.

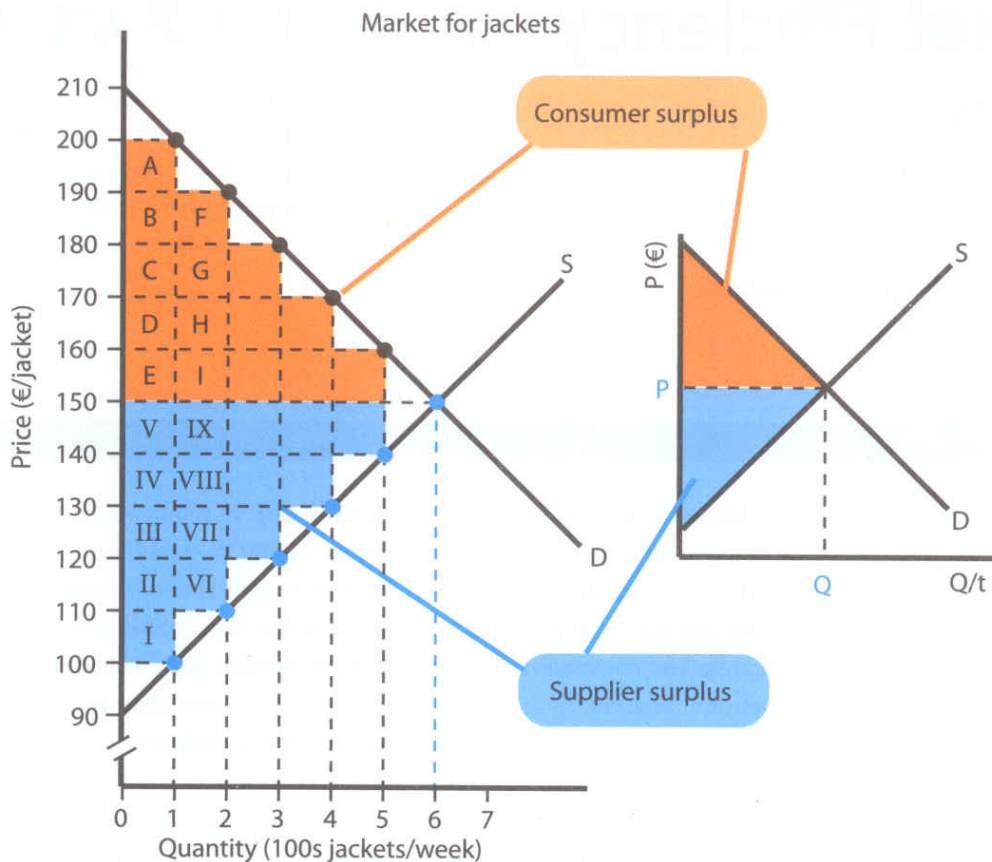2 Now think; why are post-Christmas sales so enjoyable?!

Figure 8.1: Consumer and supplier surplus

Producer/supplier surplus

Suppliers providing goods at the market price also receive a surplus; the price which is above and beyond the price at which certain suppliers would be willing (and able) to provide the good at.

- **Areas I to V**: The supply curve in Figure 8.1 shows that suppliers would be willing/able to put 100 jackets on the market at a price of €100. At the market price of €150 the revenue per jacket received is €50 above the marginal cost of producing these 100 jackets. Areas I, II, III, IV and V show the **supplier surplus**, i.e. the total additional marginal revenue above and beyond what suppliers would be willing/able to put on the market for the first 100 jackets.

- **Areas VI to X**: At an even higher price of €110, producers add an additional €40 of surplus, areas VI to IX...and so on up to the market clearing price of €150, At a market price of €150, producers receive a total supplier surplus of the blue triangle in the diagram on the right.

In summa: As with consumer surplus, if the transaction takes place, then there is a marginal benefit for the supplier.

Definition: 'Producer/supplier surplus'

Producer or **supplier surplus** is the additional benefit (revenue) received by a producer in selling a good where the market price is above what the producer would be willing to accept. (HL: this is in fact *marginal cost*. More in Chapter 23.) Total supplier surplus is the area above the supply curve and below the market price.

Allocative efficiency

The perfectly competitive market at equilibrium price will maximise both consumer and supplier surplus, leading to an **optimal allocation of resources**. Any action which reduces this **total surplus** (also known as community or societal surplus) renders an allocatively suboptimal outcome. (See for example *Monopoly*).

Maximum societal (community) surplus

If pure market forces have resulted in equilibrium in all markets, then the societal surplus will be maximised. Skim back and revise the example of the price mechanism and resource allocation in Chapter 7 (Figure 7.2). The increase in demand for red pens meant that resources were re-allocated to red pens as in Figure 8.2 below.

a) Red Pens — Increased societal surplus after resources are re-allocated to red pens.

An **increase in demand** for red pens (D_0 to D_1) increases both consumer and supplier surplus (= societal surplus)

b) Green Pens — Decreased societal surplus after resources are re-allocated to red pens.

A **decrease in supply** for red pens (S_0 to S_1) decreases both consumer and supplier surplus (= societal surplus)
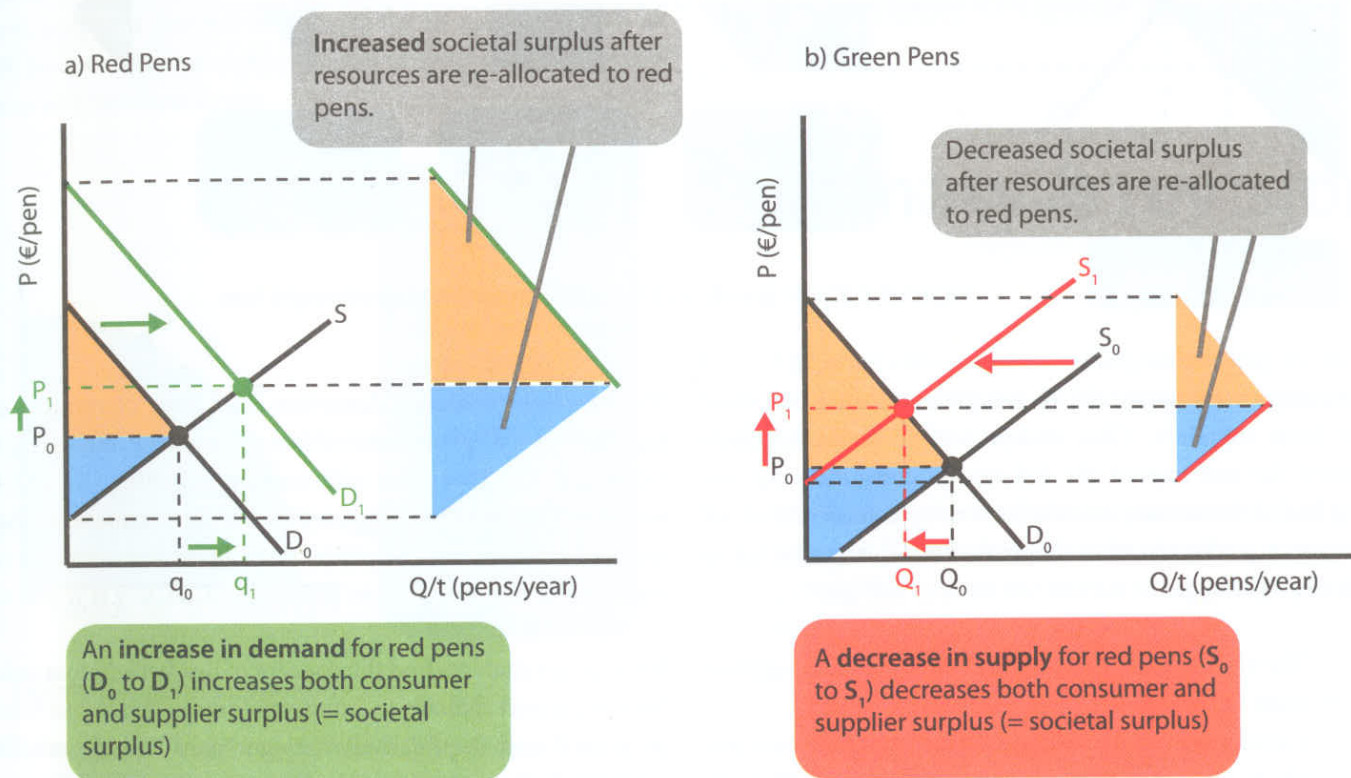
MICROECONOMICS

Figure 8.2 Re-allocation and societal surplus

The basic point here is that *both* periods are allocatively efficient! Before the demand for red pens increased, both markets were optimally efficient – there was no excess supply or excess demand. After demand for red pens increased and resources were re-allocated for green to red pens, societal surplus is *still maximised* given the limitations of scarcity in resources. Again, the question of "…which is the best *combination* of quantity of goods…" isn't the issue. The issue is how the price mechanism in a freely competitive market serves to allocate resources in such a way as to optimise use of available resources.
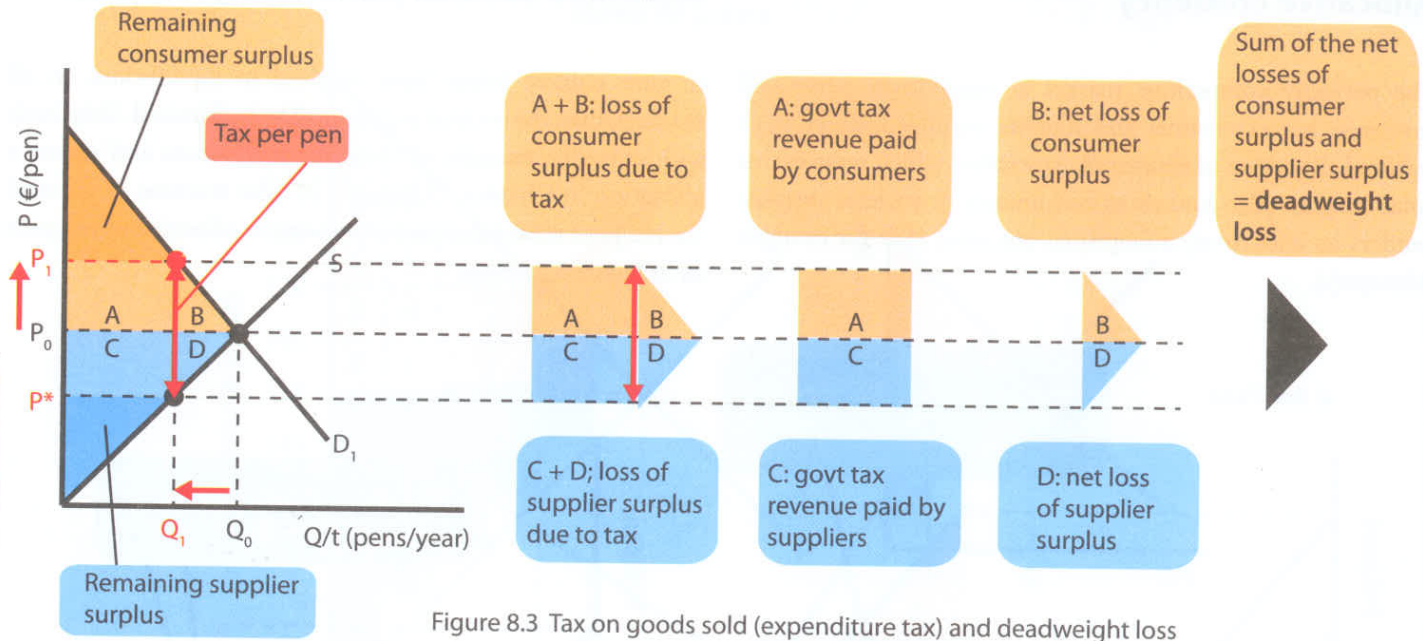
Deadweight loss

While HL will go into the issue of **deadweight loss** in Chapter 13, it is such a useful concept that I introduce the basic concept here for SL. Having said the above about maximum societal surplus when free market forces rule, it is also possible for

society to incur (= acquire) an overall – *avoidable* – loss of societal surplus. Government intervention, such as a tax on goods sold can result in lower overall societal surplus, as can imperfect competition such as monopolies.

Instead of producers voluntarily reducing the supply of green pens in order to re-allocate resources to red pens, assume that the government instead levies (= puts) a tax on pens sold, e.g. an expenditure tax. Since a tax on goods sold by firms is an indirect tax (consumers do not pay the tax to government but to suppliers who in turn pay the tax on to government) suppliers will have to decide how much of the tax they can get consumers to pay.[3] In the case outlined in **Figure 8.3**, the tax is shown by the double-edged red arrow and the change in price is from P_0 to P_1 – half the tax.

---

3   This deals with the price elasticity of demand covered in Chapter 9. Exciting, no?!

MICROECONOMICS

Remaining consumer surplus

Tax per pen

A + B: loss of consumer surplus due to tax

A: govt tax revenue paid by consumers

B: net loss of consumer surplus

Sum of the net losses of consumer surplus and supplier surplus = deadweight loss

C + D; loss of supplier surplus due to tax

C: govt tax revenue paid by suppliers

D: net loss of supplier surplus

Remaining supplier surplus

Figure 8.3 Tax on goods sold (expenditure tax) and deadweight loss

If consumers are paying half the tax then the other half must come from suppliers. (This makes perfect sense; suppliers evidently did not impose the full tax on consumers but are taking half of the tax they need to hand over to government out of their own pockets.) Let us follow the "flow" from the tax to the deadweight loss:

- The tax decreases the quantity demanded and supplied from Q_0 to Q_1
- Consumers pay more (P_0 to P_1) but suppliers earn less (P_0 to P^*) since they have to pay the tax of P^* to P_1 shown by the red double-edged arrow
- Total tax revenue to government is the tax per unit (P^* to P_1) times the quantity (Q_1) – e.g. areas **A** and **C**
- Since consumers are paying 'more for less' then there must be a loss of consumer surplus – this is shown by areas **A** and **B**
- Suppliers are receiving 'less for less' which is a loss of supplier surplus – areas **C** and **D**
- Total loss of supplier surplus and consumer surplus, societal surplus, are areas **A + B + C + D**
- However, government is part of society just like firms and consumers – this total (gross) loss of societal surplus is to an extent offset (= compensated) by the increase in government revenue of **A + C**
- The two areas of societal surplus lost which are not offset by government's gain (tax revenue) are areas **B** and **D** – this **net loss of societal surplus** is the deadweight loss due to the expenditure tax on pens

The tax creates a market distortion since both consumers and suppliers would have benefitted from consuming and selling Q_0 rather than Q_1. Every pen sold/consumed between Q_1 and Q_0 would *add to overall societal surplus* but the tax prevents these transactions from taking place. Since these lost units do not result in tax revenue going to government, society is worse off.

Many of my students find the concept of deadweight loss rather frustrating and difficult to intuitively grasp. "Yeah, but where do areas B and D go?!" Well, they go "poof". Disappear. "But WHERE?!" This can go on for a long time. Then I tell the story of losing one's wallet in the forest. If I lose a wallet with $200 in it but somebody else finds it, then society is in fact just as well off since my loss has been offset by someone else's gain. However, if I lose the wallet and $20 is eaten by hungry squirrels and then is found, my loss is not entirely offset by someone else's gain. The $20 went "poof". This is the net loss of consumer surplus.[4]

> **Definition: 'Deadweight loss'**
>
> A **deadweight loss** is an efficiency loss to society since societal surplus is decreased without a corresponding gain for another actor in society. It is the **net loss** of consumer surplus and supplier surplus.

4 "….hrrrrrrrrrrnggggggggggrrrrrr…BUT WHERE DOES THE DEADWEIGHT LOSS TRIANGLE GOOOOOOO…?!" "Look guys, it goes "POOF"! It is an ex-triangle! It is pushing up daisies! Gone to the Happy Triangle Hunting Grounds! It is no more! Extinct! Hasta la vista baby! Croaked! Stuck its spoon in the wall! Toe-tagged! Living-impaired!" I am not paid enough for this.

Pareto optimum

The concept of societal efficiency in an economy was pioneered by the Italian/French social-economist Wifredo Pareto (1848 – 1923) and is often explained using a PPF supporting illustration. For once, I shan't be different. Societal efficiency is said to be maximised when the economy is operating *on the PPF*; anything else would mean that resources are not being used to the utmost. The PPF in Figure 8.4 shows an economy where output is divided into goods for IB teachers[5] (such as TOK books and comfortable shoes) and IB students (Nintendo games and marshmallows).
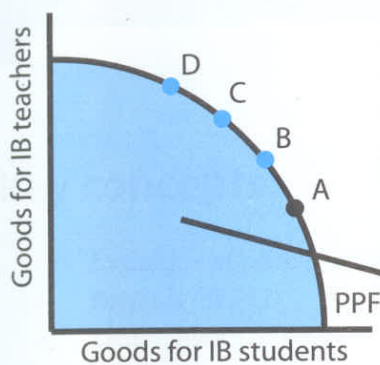
When all available resources are fully employed, increasing output of teachers' goods means taking resources away from producing students' goods. Each point on the PPF is a *Pareto optimum* point of output.

Figure. 8.4 *Pareto optimum*

At **point A**, the economy is *maximally efficient* and it is impossible to increase teachers' goods without decreasing output of students' goods – and vice versa. Therefore, anywhere along the PPF, the economy can be said to be optimally societally efficient. This is referred to as a *Pareto optimum*. The condition is put as a conditional phrase: 'If it is impossible to increase the well-being of one person [here IB teachers] without making someone else [IB students] worse off, then the economy is at a **Pareto optimum**.' Note that Pareto optimality in no way suggests the "best" point of output, only that **points A to D** fulfil the criteria for optimal efficiency.

Summary & revision

MICROECONOMICS

1. The price consumers are willing/able to pay minus the de facto market price is the **consumer surplus** for a consumer.

2. The market price a supplier receives for a unit of goods minus the minimum price the supplier demands for putting that unit on the market is **supplier** or **producer surplus**.

3. The sum of consumer and supplier surplus is **societal** (or community) **surplus**.

4. When the market is in equilibrium there is no excess in supply or demand so the use of (scarce) resources is optimised. This is **allocative efficiency**.

5. Any form of market distortion which leads to a net loss of societal surplus is a **deadweight loss** to society. This means that there is allocative **in**efficiency.

6. If all resources are used to the maximum then the economy would be on the PPF. It is thus impossible to make consumers of Good X better off without making consumers of Good Y worse off. This is known as **Pareto optimum** and is another way of defining optimal allocative efficiency e.g. when it is impossible to make one group of consumers better off without making another group worse off..

5 There are so many things I'd love to put here…but my editor won't let me. Feel free to write to me with suggestions!

MICROECONOMICS

1.2

9. Price Elasticity of Demand

Key concepts:

- Definition of price elasticity of demand (PED)
- Formula for PED
- Determinants of PED
- Applications of PED

Section 1.2 deals with the responsiveness in quantity supplied or demanded for goods due to changes in variables which influence them, such as price and income. Looking at these relationships allows us to decipher the choices and actions of the major actors on the market; firms, households and governments.

Definition of price elasticity of demand (PED)

You have no doubt already reflected upon how different goods must have different patterns of demand, i.e. different demand curves. For example one's demand for, say, food should be different from that for wrist watches.[1] It can easily be portrayed diagrammatically.

---

1 But again, people are not the same in their preferences! I remember living on oatmeal for six months as a poor student so I could by a USD2,000 Baume & Mercier watch. It bears repeating that the demand curve is the aggregate of individual demand.
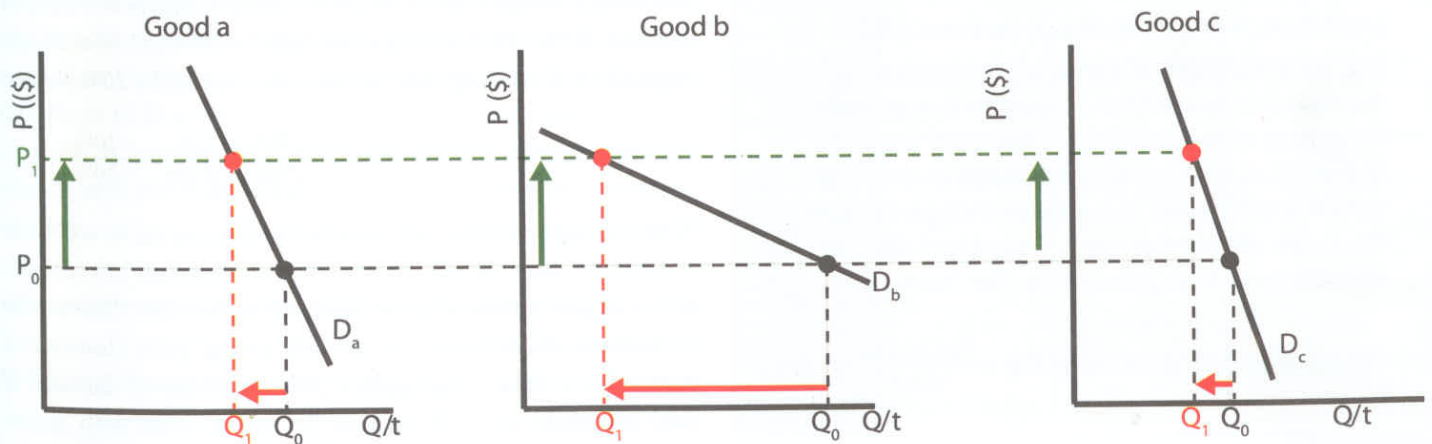

Fig. 9.1 Demand

Different patterns of demand

Let us look at a few different *untitled* demand curves in Figure 9.1 and see if you can set correct captions to them. Choose between *drugs*, *DVD films*, and *cinema*. Send me $5 for correct answers.[2]

It makes sense doesn't it? The demand curve will be very 'steep' for some goods – i.e. our quantity demanded will not decrease a great deal when the price of the good increases. (Note: using the term 'steep' in reference to a change in quantity demanded resulting from a change in price is faulty use of economic terminology. The correct term is *price elasticity of demand* as shall be seen.) Other goods will show a marked decrease in quantity demanded should the price rise; a rise in the price of DVDs will decrease quantity demanded quite considerably. DVDs are said to be *price sensitive*. At the same time, a massive increase in the price of my asthma medicine will probably have little effect on my quantity demanded.

The common denominator is often the availability of goods which we could buy instead within a given period of time; *substitute goods*. The more available substitutes there are (and the closer they are) the easier it is to switch to another good. The increase in the price of DVDs leads to a larger relative (percentage) decrease in the quantity demanded. Conversely, a rise in the price of asthma medicine, which has few substitutes, would mean that quantity demanded would not fall significantly.

Calculating price elasticities is no more complex than calculating a percentage increase in income; it is simply a matter of being stringent and consistent. Calculate the percentage change of quantity and divide this by the percentage change in price. Just remember to be consistent in your use of starting values.

> **Definition: 'Price elasticity of demand (PED)'**
>
> The price elasticity of demand is a measure of the responsiveness of the quantity demanded for a good with respect to a change in the price of the good. It is the *relative* change in quantity demanded (Q_d) due to a *relative* change in price (P). As the demand curve is downward-sloping, PED will have a negative value. The formula is:
>
> $$\text{Price elasticity of demand (PED)} = \frac{\%\Delta \text{ in Qd}}{\%\Delta \text{ P}}$$

Formula for PED

The correct term for the measure of price sensitivity in the demand for goods is price elasticity of demand (PED). In short, we are looking at how responsive demand is to an increase or decrease in the price of a good, which can be put as; 'What is the *relative* change in quantity demanded due to a *relative* change in the price?'

Putting the quote above in formulaic language, we get:

$$\frac{\%\Delta \text{ in Qd}}{\%\Delta \text{ in price}}$$

> **Exam tip; referring to elasticities in exam answers**
>
> When either using elasticities or defining them, always include the basic formulaic expression – just to make sure. Your definition should include the term relative – e.g.'…the relative change in quantity demanded due to the relative change in price…' Also, remember 'Quantity rules!' or 'Quantity on top!' – no matter which measure of elasticity you are using, quantity is always the numerator ("on top" in the formula).

Note the key word here, 'relative'. An increase in price of $5 simply does not mean the same thing when talking about a Jaguar car or a bag of crisps. Similarly, a decrease in quantity demanded of 500 units per week doesn't mean the same for Jags as for bags of crisps. We must put the change into relative terms by using *percentage* increase/decrease.

By dividing 'percentage change in quantity demanded with percentage change in price' we get a value of the price elasticity of demand. If the price of margarine **rises** by 10% and households respond by decreasing their quantity demanded by 20% we get;

$$\text{Price elasticity of demand (PED)} = \frac{\%\Delta \text{ in Qd}}{\%\Delta \text{ in price}} = \frac{-20\%}{+10\%} = -2$$

Note the *negative value*; any downward sloping curve will have a negative slope, which means that all demand curves will have a negative value of price elasticity of demand. This is why economists seldom bother in actually saying 'price elasticity of demand is minus 2' but simply 'price elasticity of demand is two.'[3] However, do NOT confuse the values above with 'slope'. The value of –2 is just the value of price elasticity of demand

---

2 No, you'd better not. The answers are; Figure a = cinema, Figure b = DVDs, Figure c = drugs. Note that drugs can be pharmaceuticals as well as the Mexic…that's, illegal variety.

3 The mathematically inclined will realise that we are simply referring to PED as an absolute value, i.e. |2|

along a given section of the demand curve (or, more commonly, at a given *point* on the demand curve), nothing else. Nor do price elasticity values have any sort of relativistic meaning, such as kilograms and centimetres might have. What I mean by this is that while two kilos is twice as heavy as one kilo, a price elasticity of 2 is not 'twice as elastic' as a price elasticity of 1.

DANGER MINES WARNING !

Elasticities

Once again I should warn you. An all-too-often made mistake in using the concept of elasticity is to confuse the value of elasticity with the slope of the curve. Any straight-line curve will have a slope that is constant – while the value of PED along the same slope will vary.

I also strongly advise against the use of any alternative formula. There are a number of ways to re-write the basic formula of 'percentage change in quantity over percentage change in price'. The exam halls are littered with the bodies of students who have tried to use these alternatives and monumentally failed.

A Little Depth

Using percentages as a measurement of an increase or decrease can be a bit confusing. If my income goes from £100 to £125 (+ £25) then I have increased my income by 25%. If my income falls from £125 to £100 (- £25) then my income has decreased by 20%. A percentage change is arrived at by dividing the change with the original value. In formula, where 'y' is personal income:

Increasing my income from $100 to $125

Percentage change in income $(y) = \frac{y_1 - y_0}{y_0} \times 100$

$$\rightarrow \frac{125 - 100}{100} \times 100 = +25\%$$

Decreasing my income from $125 to $100

Percentage change in income $(y) = \frac{y_1 - y_0}{y_0} \times 100$

$$\rightarrow \frac{100 - 125}{125} \times 100 = -20\%$$

In all formulae I will use, all variables using the lower '0' signify the starting value and '1' the next value. You have already seen this in supply and demand curves. In the formulae above, y0 is the starting value; $100 in the first case which increases to$125. In the second case, y_0 is $125 and y_1 is $100 as we have lowered income. As long as one is consistent in the use of the starting values in this formula it is impossible to go wrong.

Range of PED values

Let us use three different goods, imaginatively re-using the goods from Figure 9.1; cinema, DVDs and pharmaceutical drugs. The original quantity demanded is 100 for each good and all three have an initial price of 10. (Using different currency units in order to be a bit realistic.) Now *we raise the price to 12*, an increase in price of 20%, causing the quantity demanded to fall by a certain amount for each good.

Here's a simple table of the outcomes in terms of quantity demanded when the price is raised from 10 to 12, and the calculated PED for each good;

Good	Cinema	DVDs	Drugs
$P_0 \rightarrow P_1$	€10 → €12	$10 → $12	£10 → £12
ΔP	+€2	+$2	+ £2
%ΔP	+20%	+20%	+20%
$Q_0 \rightarrow Q_1$	100 → 80	100,000 → 40,000	100,000 → 90,000
ΔQd	−20	−60,000	−10,000
%ΔQd	−20%	−60%	−10%
PED	−1.0	−3	−0.5

MICROECONOMICS

The PED values at the far right were calculated by using the formula % ΔQd / %ΔP. Here is the PED calculation for **cinema**[4]:

$$PED_{cinema} = \frac{\%\Delta \text{ in Qd}}{\%\Delta \text{ in price}}$$

$$= \frac{\frac{Q_1 - Q_0}{Q_0} \times 100}{\frac{P_1 - P_0}{P_0} \times 100}$$

$$= \frac{\frac{80 - 100}{100} \times 100}{\frac{12 - 10}{10} \times 100}$$

$$= \frac{-20\%}{+20\%}$$

$$= -1.0$$

In the example, all three goods' prices were raised by the same *relative* amount; +20%. The percentage decrease in quantity demanded for each good was different: cinema –20%; DVDs –60%; and pharmaceutical drugs –10%. This gives us a range of values:

- The least price sensitive (**price inelastic**) good is drugs, where a 20% increase in price resulted in a modest 10% decrease in quantity demanded, e.g. **PED = -0.5.**
- The most price sensitive (**price elastic**) good is DVDs, where the same 20% increase in price led to a 60% fall in quantity demanded; **PED = -3.**
- Cinema sales fell by the same proportional amount (**unit elastic**) as the rise in price; **PED = -1.** Please keep reminding yourself that throughout the discussion on elasticities the subject of *absolute* changes in price or quantity demanded is largely irrelevant – we are only interested in the *relative* changes.

> **Definition: 'Range of PED values'**
>
> A price elasticity of demand of less than 1 is called inelastic demand; PED that is equal to one is unit elastic; PED that is higher than 1 is elastic. (Note that I have dropped the minus sign, e.g. used absolute values. |x|)
>
> PED < |1|; inelastic   PED = |1|; unit elastic   PED > |1|; elastic

---

4   Cinema tickets have an estimated PED of |0.9|.

PED along a linear demand curve

Economists tend to address most questions with 'It depends…' or 'Assume that…'. This is unfortunately true even in the case of calculating price elasticities. In moving along the curve and plotting out the points, one will get two different values, depending on whether the price falls or rises.

In Figure 9.2 following, the detail of the PED for DVDs is shown. In raising the price from \$10 to \$12, the quantity demanded fell from 100,000 to 40,000, giving a price elasticity of demand of 3. (Note that I have dropped the minus sign now.)

However, if the initial price is at \$12 and it *falls* by \$2 to \$10, we get:

In lowering the price we got a higher value of PED – even though we used the same figures. This is the **arc** elasticity of demand which shows the value of PED over the arc of two points. The way to countermand this problem is to calculate a *mid-point value* by using average values of price and quantity, which is done in 'Outside the box' following as it is not directly part of the syllabus. (Do not be alarmed by the seeming complexity of the formula as you will never be required to use it in actual calculations.)
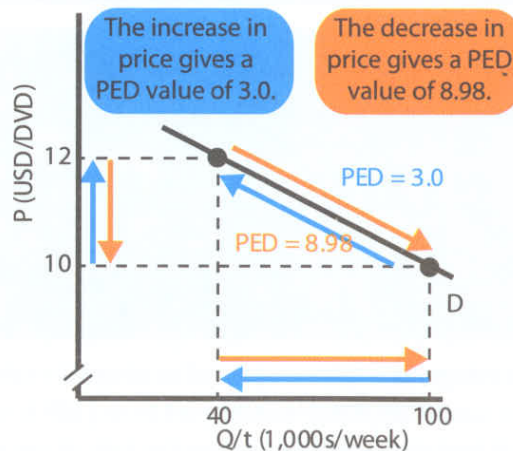

Figure 9.2 Detail of differing PED for DVDs

I have deliberated long and hard as to the best way of showing price elasticity values and have finally arrived at the following; all references to PED will be based on the values at a *single point* unless otherwise stated.[5] Feel free to skip the next 'Outside the box' and simply move on to the section on elasticity along the entire demand curve. Just keep in mind that PED values used in this book are in fact based upon the mid-point formula unless clearly stated or otherwise shown.

---

5   I went crying to my friend and colleague, Joe the math teacher. He simply said 'This elasticity stuff is a bunch of nonsense.'

OUTSIDE THE BOX

The computational paradox of using the arc PED formula can be solved by using average changes in the price and quantity. This is the mid-point method, which gives us a value of PED that is in between the two different values we would get by using the arc formula given earlier.

The values we got before were:

Raising the price from 10 to 12:

$$= \frac{\frac{40 - 100}{100} \times 100}{\frac{12 - 10}{10} \times 100} \rightarrow \frac{-60\%}{+20\%} = 3.0$$

Lowering the price from 12 to 10:

$$= \frac{\frac{100 - 40}{40} \times 100}{\frac{10 - 12}{12} \times 100} \rightarrow \frac{+150\%}{-16.7\%} = -8.98$$

The mid-point formula uses the same change in price and quantity as before – but uses an average value as the denominator. Thus we get the same PED no matter whether we are moving up or down the demand curve.

Raising the price from 10 to 12:

$$\frac{\left|\frac{Q_1 - Q_0}{\left(\frac{Q_0 + Q_1}{2}\right)}\right| \times 100}{\left|\frac{P_1 - P_0}{\left(\frac{P_0 + P_1}{2}\right)}\right| \times 100} \rightarrow \frac{\left|\frac{40 - 100}{\left(\frac{100 + 40}{2}\right)}\right| \times 100}{\left|\frac{12 - 10}{\left(\frac{10 + 12}{2}\right)}\right| \times 100} \rightarrow \frac{-85.7\%}{+18.2\%} = -4.71$$

Lowering price from 12 to 10:

$$\frac{\left|\frac{Q_1 - Q_0}{\left(\frac{Q_0 + Q_1}{2}\right)}\right| \times 100}{\left|\frac{P_1 - P_0}{\left(\frac{P_0 + P_1}{2}\right)}\right| \times 100} \rightarrow \frac{\left|\frac{100 - 40}{\left(\frac{40 + 100}{2}\right)}\right| \times 100}{\left|\frac{10 - 12}{\left(\frac{10 + 12}{2}\right)}\right| \times 100} \rightarrow \frac{+85.7\%}{-18.2\%} = -4.71$$

The price elasticity of demand at the price of USD11 (the mid-point between USD10 and USD12) is 4.71. This value lies between the two arc PED values of 3.0 and 8.98.
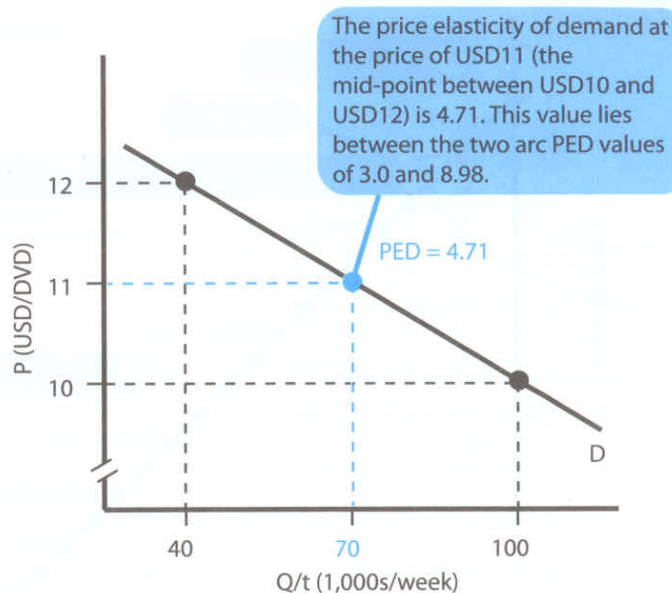
PED = 4.71

Figure 9.3

PED along the demand curve

Elasticity varies not only between different goods, but also along the demand curve for any specific good portrayed using a straight-line demand curve. A good 'law' to keep in mind when dealing with PED for straight-line demand curves is that the *price elasticity of demand will go from infinity to zero with unit elasticity in the middle.*

I now return to the demand for DVDs. The demand curve in Figure 9.4 shows the quantity demanded for DVDs in thousands of films per week at specific points along the curve. I have used these values to calculate the (point) PED values (blue dots) along the curve – and since all demand curves are downward sloping, we again disregard the minus sign. The diagram illustrates how PED values change along the slope of a straight-line curve.

There are three notable points along the demand curve in Figure 9.4.

- The first shows that the value of **PED on the P-axis is infinity** ('A' in the figure.) Pure mathematics says that anytime one tries to divide a value by zero the result is undefined (or, less correctly, infinite). At **point A** we would be dividing a change in quantity with an average price change of *zero*, thus we have a PED of infinity, the sign for which is ∞.
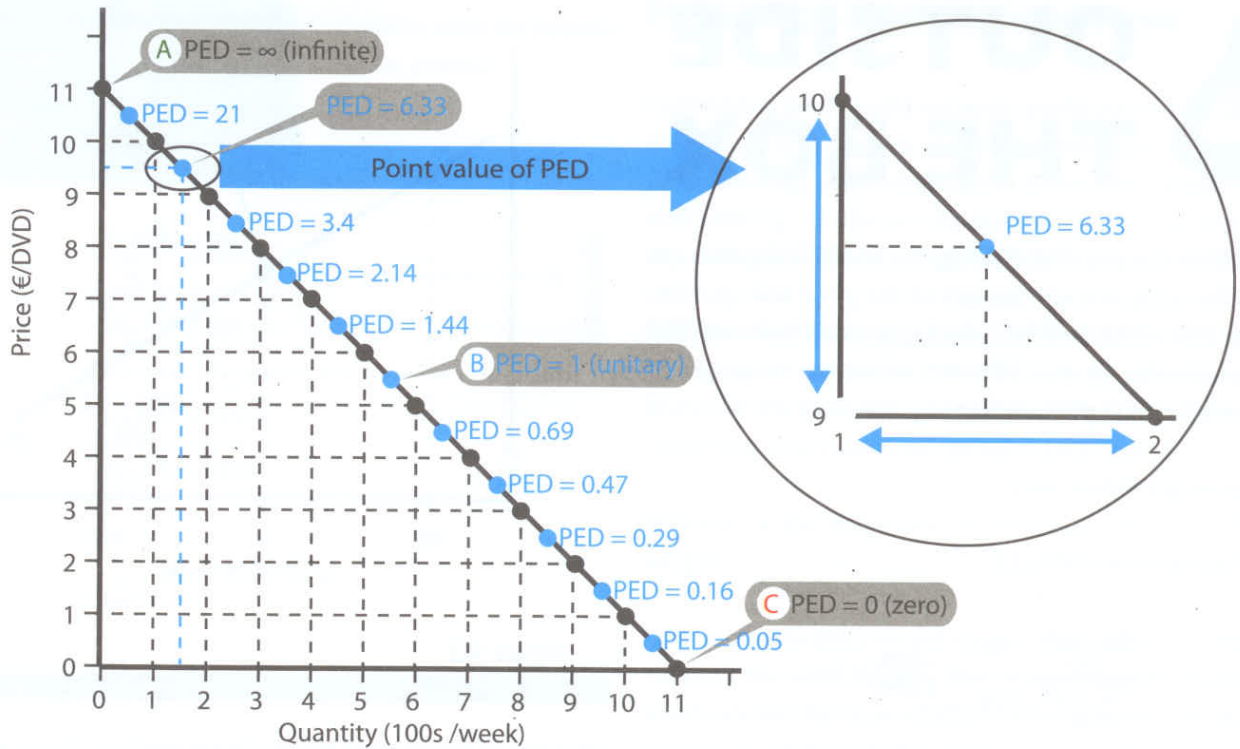
Fig. 9.4 Demand for DVDs – values of PED along a linear demand curve

- Moving downwards from the point of infinite elasticity, PED falls consistently until a point right in the middle of the curve, **point B**, at a price of 5.5 and 5,500 in quantity demanded. Here the **PED is 1**, or **unitary**. Unitary PED means that the quantity demanded changes by the same proportion as price, for example an increase in price of 10% would decrease quantity demanded by 10%. At any point below the price at unit elasticity, PED is lower than 1.

- At **point C** where the demand curve intercepts the Q-axis we have **zero price elasticity of demand**. PED is zero here as the numerator in the formula for PED is zero, since the average percentage change in quantity is zero and whatever denominator will result in an elasticity of zero, no matter what the change in price.

Figure. 9.5 PED along a linear curve

Three 'extremes' of PED

There are three additional cases which together cover all three 'extremes' of price elasticity of demand. These are shown below in Figure 9.6a, b and c.

- Figure a shows a good which has zero price elasticity, i.e. where the same quantity is demanded no matter what the price.

- Figure b shows the opposite; quantity demanded is infinite at a price of four or below (but zero at any price above $4).

- Figure c shows a good which will have the same proportional response to any change in price – PED is unit elastic at all price levels. Any proportional change in price will lead to the same proportional change in quantity demanded. Note that no matter what the price, P × Q yields the same value of total revenue.
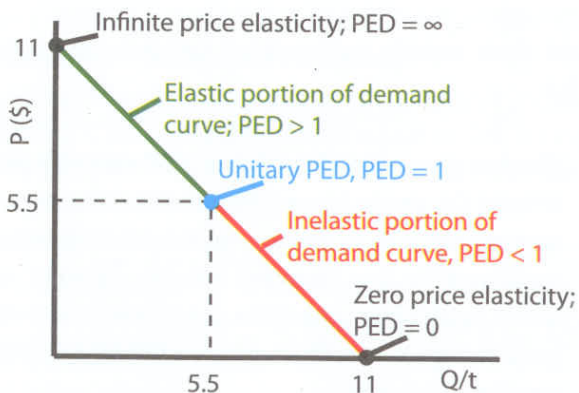
All three demand curves have a single value of PED. A vertical demand curve will be totally price inelastic (PED = 0) while a horizontal curve will be infinitely price elastic (PED = ∞). A curve which has the same area under the curve no matter what point one is on, will have unitary elasticity (PED = 1) all along the curve.[6]

---

6   (Such a curve, according to Joe the math teacher, is called a 'rectangular hyperbola'. Now go and win a game of Trivial Pursuit. Who makes these names up, anyhow?! Rectangle?!)
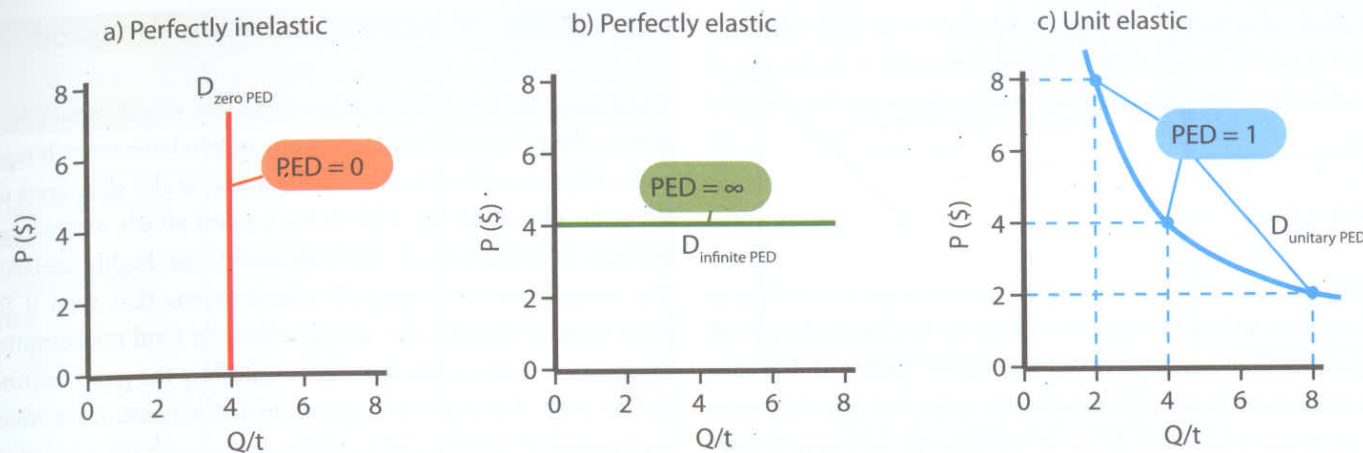
Figure 9.6 a), b) and c)

Determinants of PED

There are any number of factors that will have an influence on the price elasticity of demand, all of which deal directly or indirectly with the strength of consumers' preferences and their ability to find/accept other goods instead. One realises intuitively that the elasticity of demand for medicines will differ from that of luxury cruises in the Caribbean. The determinants of price elasticity of demand can be outlined in three headings;

1. The availability/closeness of *substitutes*;

2. The *time span* involved; and

3. The *proportion of income* spent on the good.

Availability and closeness of substitutes

PED basically measures how responsive consumers are to a change in the price of a good, and the closeness/availability of substitutes is doubtlessly the main determinant of price elasticity of demand. Just like the moving along a PPF induces choice and giving something up, so too does switching from one good to a comparable good. The choice is based on the **ability** and **willingness** of consumers to give up Good X for Good Y – both of which are determined by whether there is in fact a substitute in the first place and secondly whether there is a high degree of perceived similarity, or closeness in the two goods. The rule of thumb is that the closer the perceived substitute, the higher the value of PED, i.e. *high substitutability* leads to *greater price elasticity* of demand.

Notice that I use 'perceived' closeness. This is not by accident. While milk and cream are chemically quite similar, I know of few people who would substitute milk with cream as a refreshing drink – and I certainly wouldn't want to have a seat next to them on an airplane. At the same time, Rolls Royce

cars and yachts are highly dissimilar, but are in fact relatively close substitutes! Beauty is in the eye of the beholder, and so too is substitutability. The *propensity* and *willingness* of people to substitute one good for another defines whether goods are close or not.

Many goods are conceivably not substitutable at all. My asthma medicine, cigars and compressed air for scuba diving would be considered highly inelastic goods….by ME… and, OK, by most others too.[7] Again, it is the aggregate of individual choice that builds market demand and therefore price elasticity. Alcohol, tobacco and other mind-altering substances would all have low values of PED as they would be non-substitutable for many people's habits or addictions.

It all comes down to the perception of people and also how one *defines markets*. The wider the definition of goods, the lower the value of elasticity – as the wider definition would include any possible substitutes. For example, my addiction to coffee is shared by a good number of people; coffee has an estimated PED of 0.25 (highly inelastic) but if we include a number of coffee substitutes in our definition of the market – say tea and cocoa – and call the market 'hot beverages' then elasticity would fall even further. In including possible substitutes in the grouping, we further limit the scope of possible substitution.

Highly substitutable goods would have many examples within the rubric (= heading) of *brands*. In Greece I always buy the local beer, Mithos, instead of the locally brewed Amstel/Heineken. Is it a preference? No, I couldn't tell the difference to save my life. However, Mithos is always a little cheaper and as they are perfect substitutes…. Perhaps now it makes sense

---

7   It just struck me; I have asthma and live in one of the most polluted cities in the world, Jakarta, smoke cigars, run Marathons, scuba dive.

that what advertising and marketing does is to help convince us that Brand X is superior to Brand Y. Beauty is in the eye of the beholder – but values change and perhaps can be changed for us.

Time span

Demand tends to be more price inelastic in the short run than in the long run. When I bring this subject up in class, I always ask my people if anyone took a taxi during the weekend. I always get a few raised hands and it inevitably turns out that they were at a party and took a cab home at 03:00 in the morning.8 Taxis are notoriously expensive – plus the fact that the night rate is higher – yet none of the kids even considered alternatives to the taxi. This is an example of how a short time span will lead to low elasticity of demand; there are simply fewer viable options within a short span of time than within a longer period.

When the first oil crisis hit in 1973/'74 and the price of oil quadrupled, the demand for oil was virtually perfectly price inelastic in northern countries, as sub-zero temperatures were the case during the crisis and there simply were no credible alternatives in the time span of those first winter months. Come spring, the first thing people did was to increase insulation, install double-glazed windows, build wood-burning fire places, and install thermal energy pumps….everything to decrease dependency on oil. This meant that over a longer period of time, the PED for oil increased as people had time to set up viable options and adjust.

There are also a goodly number of goods which one can postpone the purchase of, such as sofas and interior decorations, while other goods are considered necessities, such as electricity and, em, well, cigars.9 I often say that there is, in fact, only one determinant of demand; the availability of substitutes. 'Time' is just another way of saying that one will have the incentive, opportunity and propensity to search for viable substitutes. Short time-span → lack of adjustment and search time for substitutes → inelastic demand.

Proportion of income spent on the good

'How many of you bought paper clips this week? None? How about rubber bands? No? How many of you have *never* bought paper clips or rubber bands?' Every hand in the class goes up. What do you think the PED is for a good which represents a fraction of a fraction of one's income?! Yes, highly inelastic. The overall effect is negligible which means that even if the price were to double, the overall effect on total consumption capability will be so low that one would buy the good anyhow. In line with this argument, goods which account for a major proportion of income will be more elastic as an increase in the price would force households to cut back on outlays to make ends meet. Food and clothing are standard examples of such goods. Keep in mind though, that different income groups will have different price elasticities for goods. High income groups and high income countries will spend proportionately less of income on housing and clothing and therefore have lower PED for these goods than low income groups and low income countries.

Real life elasticities

Let's finish this discourse on price elasticities by looking at a few real-world examples of PED and putting them into the context of the above three determinants of elasticity. The table below gives the elasticity of demand for a number of goods. The figures seem to conform to what has been said about the prime determinants of price elasticity. I have number-coded the goods according to the three determinants above. Note that many of the goods overlap in terms of which determinants shape PED. For example, salt is price inelastic due to both lack of substitutes and percentage of income spent, giving it 1 and 3.

Goods	Estimated Price Elasticity of Demand	Determinant (-s)
Inelastic		
Salt	0.1	1 & 3
Gasoline, short-run	0.2	1 & 2
Gasoline, long-run	0.7	1
Coffee	0.25	1
Tobacco products, short-run	0.45	1 & 2
Approximately Unitary Elasticity		

---

8   Yeah, I tried this in my IB2 class in Mexico. The ladies looked at me with utter disbelief: "A taxi?! At 03:00? Are you out of your cotton-pickin' Swedish mind?!" It turns out that every single female student is banned by parents from taking taxis at night. It's simply too dangerous.

9   It's my book and I'll define things as I see 'em.

Goods	Estimated Price Elasticity of Demand	Determinant (-s)
Salt	0.1	1 & 3
Gasoline, short-run	0.2	1 & 2
Gasoline, long-run	0.7	1
Coffee	0.25	1
Tobacco products, short-run	0.45	1 & 2

Source: *Economics: Private and Public Choice*, James D. Gwartney and Richard L. Stroup, eighth edition 1997, seventh edition 1995.
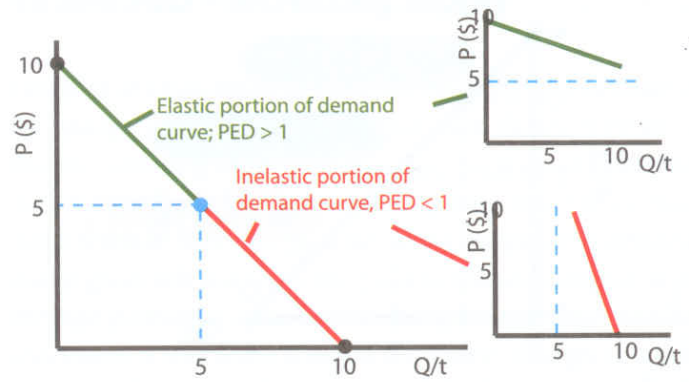

Figure 9.7 'Elastic' and 'inelastic' curves

OUTSIDE THE BOX

PED and the missing link

My people are on the ball. When dealing with PED and the issue of not confusing the value of PED with the slope of a straight line, I am often asked how one is able to refer to a straight line demand curve as being 'elastic' or 'inelastic', seeing as how the PED will go from infinity where it cuts the Y-axis to zero where it cuts the X-axis.

The answer is given in the series below. If the entire range of any given (straight-line) demand curve is portrayed in the diagram, PED will indeed vary from infinity to zero. However, when illustrating a good for which PED is highly inelastic or highly elastic, the diagram will often show only the relevant portion of the demand curve. I have broken these portions out in the two diagrams in Figure 9.7 below.

Applications of PED

We end this chapter by looking at some of the applications of price elasticity of demand. Price elasticities can provide firms with decision-making data on price and output, and governments with information on the effects of subsidies and taxes. On the international arena, PED and yED will have major implications within trade and development.

PED and total revenue (TR)

First off is the effect that sales of goods have on firms; revenue. Any child running a lemonade stand on the corner will quickly realise that sales bring in money – this is more accurately known as *revenue*. At the end of the day, when our budding entrepreneur counts total cash, i.e. total $Q_{lemonade}$ sold times the $P_{lemonade}$, he/she will have counted the *total revenue*, **TR**.

Definition: 'Revenue and total revenue (TR)'
Revenue is money earned by a firm's business activity. The total income from business activity during a time period, price times quantity sold, is the total revenue, TR. (Warning; revenue is NOT the same as profit! Profit is defined as total revenue minus total costs.)

Total revenue is easily shown in a diagram. It is the area under any specific point on the demand curve given by the price times quantity.
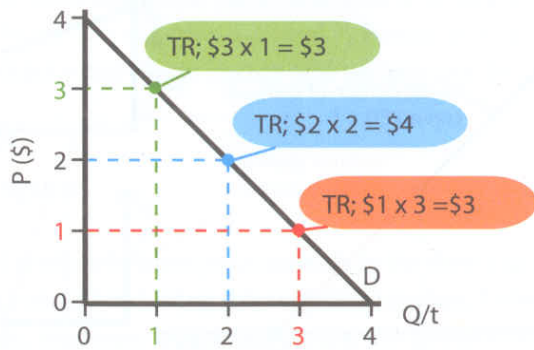
Figure 9.8 Total revenue (TR) along a demand curve

revenue, and plotting out all the various prices and quantity demanded, she soon arrives at a price of $1.25 per 25 cl glass = **$5 per litre**. (Not having finished Econ 101, she disregards any notion of profit – as will you for the time being!) At any other price, her total revenue will fall.[10]

The top diagram in **Figure 9.9** is the standard-issue S&D diagram, showing a price level from $10 to zero and a quantity from zero to 10 litres. At every given price the total revenue (TR) is shown in the total revenue diagram directly underneath the S&D diagram. At $10 per litre, quantity demanded is zero. By lowering the price quantity, demand increases as does TR – up to the price of $5. Below the price of $5 TR starts to fall, becoming zero again when the price is zero and the quantity demanded is (perhaps somewhat unrealistically) ten litres.

So, Lina sets up her lemonade stand – having done her homework. No, not geography or history – she has done a market survey in order to find out demand for lemonade. Her street corner is busy enough to have the demand curve you see below in **Figure 9.9**. She is interested in maximising her

10 When she adds gin, demand shifts to the right. Until the cops arrive.
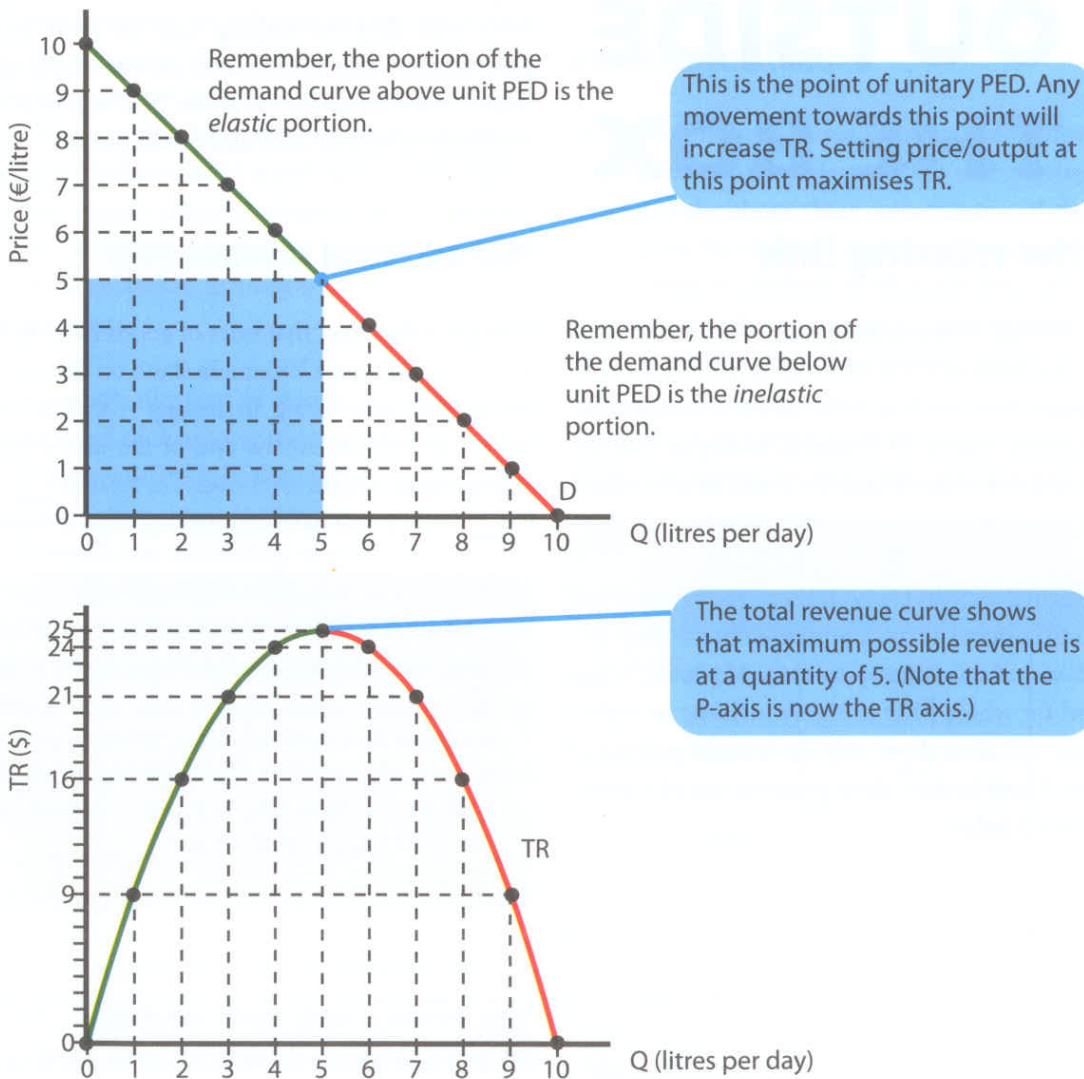

Figure 9.9 Total revenue for Lina's Lemonade Emporium

TR and PED – increasing supply

The conclusion is that *total revenue is maximised* at the point where *PED is unitary*. In other words, no matter what the initial price is, moving towards unitary PED will increase TR. One can formulate a rule for maximising TR as; 'Lowering the price for a price *elastic* good will increase TR, while raising the price for a price *inelastic* good will increase TR'.

Let's look at a two different demand curves in order to illustrate the change in TR when supply increases. Figure 9.10 shows that in increasing supply and lowering the price of a demand-inelastic good, the loss of revenue is greater than the increase. Total revenue falls. An increase in the supply of a demand-elastic good will not lower the price to the same degree as the increase in revenue – thereby increasing total revenue. Another conclusion is that when demand is inelastic, changes in supply will create large price fluctuations, while changes in supply for demand-elastic goods will cause quantity fluctuations.
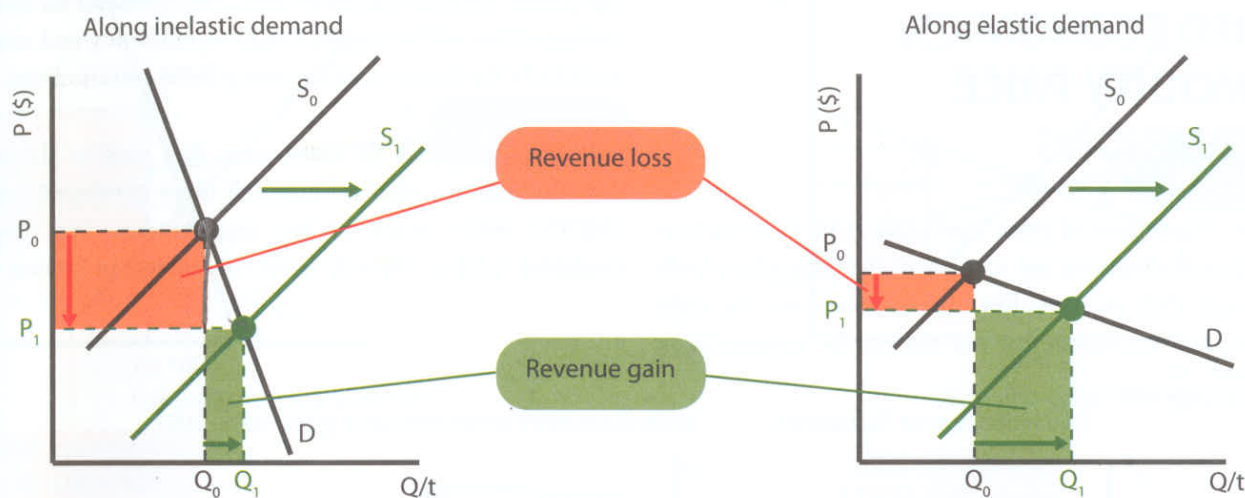
Figure. 9.10 Effect on TR due to a shift in the supply curve

PED and primary/secondary goods

Primary goods are goods which are 'raw' as in not being highly processed, i.e. have low value-added. Natural resources such as timber, iron ore and Greek marble provide raw materials for industries. Other primary goods are agricultural goods such as coffee and wheat. Many of these goods make up what is known as the commodities market.

Secondary goods are basically manufactured goods, i.e. the result of processing primary goods into consumer goods and capital, to use the widest possible collective terms. Secondary goods add value to primary goods by taking primary inputs and using labour and capital to create something more.

Tertiary goods are intangible (= non-touchable) goods, i.e. services. When you get charged a commission for changing currencies at the foreign exchange office, you are in fact buying a service. Doctors, teachers, cleaning personnel, used car salesmen and TV newscasters all provide services, i.e. tertiary goods.

Primary goods such as oil (and gasoline) and bananas tend to be relatively price inelastic in demand. There are two reasons for this. The first is that there are often no close substitutes, for example oil. The second is that such a small proportion of income is spent on the good that a large proportional change in price will not affect demand to a great extent – this would be bananas. There are also goods – salt for example – that overlap both the above, being both irreplaceable and comprising a very small amount of incomes.

Secondary goods are relatively more elastic. The main reason appears to be that there is greater substitutability. I say 'appears' with some caution since a great deal of the figures available for secondary goods will vary with how *narrowly* the market for the good is defined. While cars in general have a PED of 0.2 in the US, the PED for Chevrolet cars is 4.0[11]. This means that increases in supply for most consumer goods will increase quantity demanded substantially more than is the case for primary goods.

---

11 Source: *Economics: Private and Public Choice*, James D. Gwartney and Richard L. Stroup, eighth edition 1997, seventh edition 1995

What has been happening over the past 50 or so years, is that primary products have become ever cheaper in real terms (and in many cases even in nominal terms) and also an increasingly smaller part of total output measured in money terms. For example, the 24 major raw materials accounting for over 95% of total global output of raw materials, only accounts for 1% of global GDP.[12] Agriculture accounts for 3% of GDP in developed countries and 14% in less developed countries. The explanation for this lies primarily in increasing efficiency in production which increases supply – and since price elasticity of demand for primary goods is inelastic the price falls.

APPLIED ECONOMICS; COMMODITY PRICE FLUCTUATIONS

One of the clearer uses of price elasticities is in showing how massive price fluctuations are caused on markets where both the PED and PES are very low. The diagram on the right shows the price fluctuation on the market for commodities;
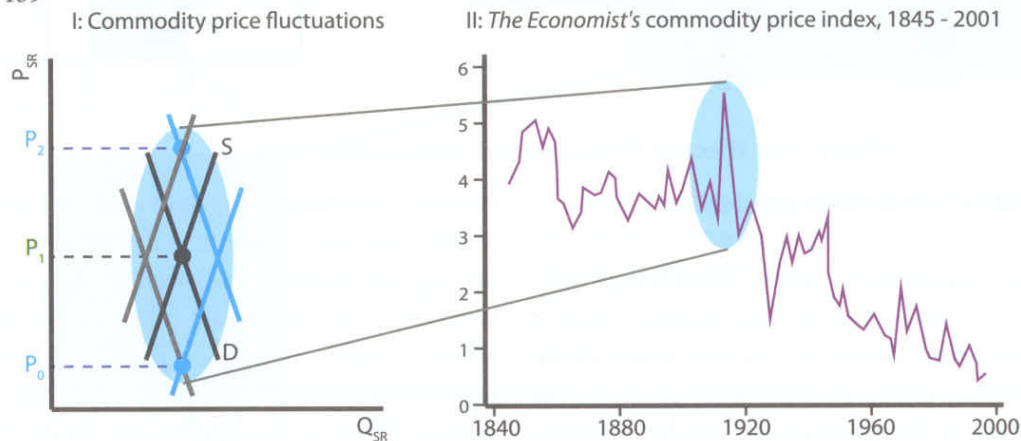
12 *Lomborg*, page 139

the Economist's industrial price index of 2000. Prices have fluctuated wildly over the past 160 years yet with a distinct downward trend. Figure 9.11 shows how low PED for these goods together with low PES causes these extreme price fluctuations.

The supply and demand curves are shown in a rather exaggerated way, but for good reason. Any good which has inelastic supply and demand will fluctuate in price more than goods which are elastic in demand. The shifts in supply for industrial commodities could be the result of seasonal variations, changes in labour and productivity, supply disruptions such as numerous wars during the period…etc. Demand for industrial commodities will be highly linked to efficiency and economic activity in the firms/countries using these commodities.

It has been the case for many years that most of the users of commodities are what we now call more developed countries (MDCs) while many of the suppliers are less developed countries (LDCs). We will revisit this subject in Section 4.

I: Commodity price fluctuations　　II: *The Economist's* commodity price index, 1845 - 2001

Figure 9.11: Commodity price fluctuations

PED and taxation

"What this country needs is a good five-cent cigar."
Thomas Riley Marshall, US Vice President 1913-1921

When I introduce this section in class, I often refer to myself as "The World's Best Tax-Payer". It's not that I don't try to avoid tax as much as the next person, I just happen to be a consumer of the type of goods favoured by governments for expenditure tax. On my desk here in my office there is a box of cigars (Cuban – thanks Ana!), a wide variety of alcohol (thanks Shota, Ji In and Daniel!) and assorted fruits and nuts (which is not my IB1 class but bananas, almonds and raisins – which I actually bought myself). OK, most of these are gifts – maybe I'm not

such a great taxpayer. Anyhow, while I pour myself a drink and disconnect the school's smoke detector so I can light up, spot the heavily taxed goods…[13]

Yes, I know; it isn't exactly rocket surgery. Governments tax alcohol and cigars far more than almonds. Why? There are two reasons:

1. The initial, perhaps somewhat intuitive answer is that smoking and drinking is harmful to me and by extension, to society in general. My use of tobacco/

13 Anyone who thinks I'm actually serious here should be sent to forced re-education clown camp.

alcohol not only damages *my* health but affects society by inflicting economic and social costs on my fellow citizens. Increased sick days, increased health care, and a possible serious case of premature death will all have a negative effect on my contribution to and burden on society. Therefore, government imposes a *disincentive* on certain goods – **demerit goods** – such as tobacco and alcohol by raising the price of such goods via expenditure taxes.

2. There is, however, an additional answer to why we tax tobacco/alcohol and not almonds; the difference in *government tax revenue* will be much higher for goods which have low price elasticities of demand than for goods with high PED.
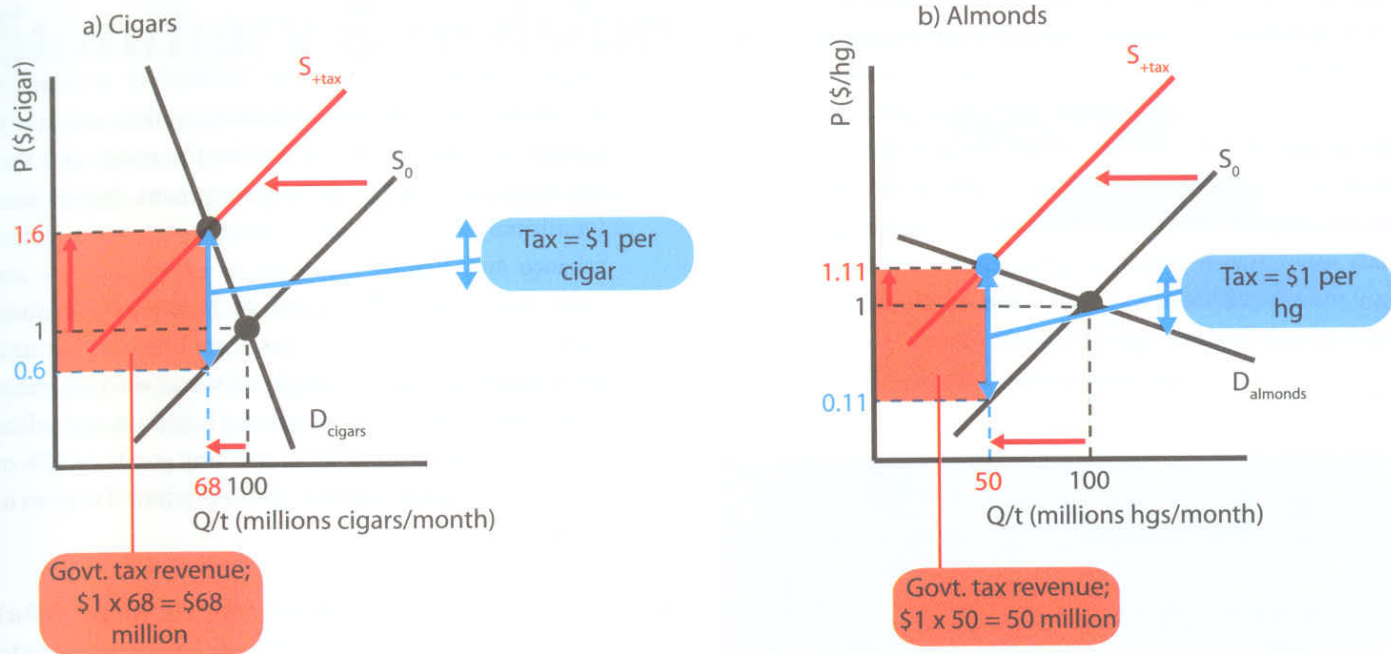
Figure 9.12 Expenditure tax on a) Cigars & b) Almonds

Note: We will re-do the issue of indirect taxation (e.g. expenditure tax) in Chapter 13.

I am now skirting the edges of HL concepts, but I hope the basic issue will come across in Figure 9.12a and **b** above. Let us assume that government wants to put a $1 tax on either each cigar or each hectogram ("hg" = 1/10 of a kg) of almonds and (for the sake of comparison) that both goods initially have the same price and quantity demanded. An expenditure tax is basically an increase in costs to suppliers and will therefore decrease supply from S_0 to S_{+tax}. This raises the price of both cigars and almonds but the quantity demand decreases by different amounts due to differing price elasticities of demand for cigars and almonds.

- Government tax revenue and tax on goods with low PED (Figure 9.12a): Assume that the original price of a cigar is $1, and the original quantity demanded is 100 million. The inelastic demand for cigars results in a fall in quantity demanded of 32% and a rise in price of 80% - which is in line with PED for tobacco of 0.4. As the

government will get $1 per cigar sold, total tax revenue is $1 times 68 million cigars; 68 million dollars.

- **Government tax revenue and tax on goods with high PED (Figure 9.12b):** Looking at taxing almond consumption by the *same proportional amount* per unit; quantity demanded has fallen by 50% and the price has risen by 10%, which is in line with the estimated PED for dried fruits of 4.6. Total tax revenue to government is $50 million – $18 million less than cigars would give.

Taxation of goods is fairly comprehensible. Goods which are either highly addictive, habit-forming or totally lacking in close substitutes are highly price inelastic and will most likely be targeted for taxation. Goods which have all three attributes will have gold and marble statues erected in their honour outside government tax offices.

Allow me to put forward a slice of personal rumination on what increasing numbers of economists are starting to question on the expenditure tax issue. One of the basic arguments for taxing tobacco is the health argument brought up above. Recent studies point to smokers actually paying a great deal more in tax than the actual cost of extra health care needed by smokers. In addition to this, a good many recent studies point out 'junk food' and obesity as a far greater health issue and economic cost to society that tobacco. Obesity related costs in America are in the area of $46 billion per year. The cost adds an estimated 36% to spending on patients and a 77% increase in medication – compared to the additional costs of smoking which are estimated at 21% and 28% increases.[14] Similarly, obesity related costs in Britain are in the area of £2.1 billion per year.[15] This is becoming ever more of an issue as rich countries around the world start to compare costs. Apart from the Norwegian government which was seriously considering raising indirect taxes on high-sugar and high-fat foods in 2007, to date, no 'obesity tax' has been levied on junk foods or high fat foods in general.

A Little Depth

Life is tax and then you die – and pay funeral tax![16] Tax revenues account for the brunt of government income and there are a number of different ways to base taxes. Taxes are mostly based on different types of economic activity and are commonly divided into two main groups.

1. **Direct taxes**: A direct tax is clearly visible to whomever is paying it; an 'in your face' tax, if you will. *Income tax* is the slice taken out of your wages/salary – for the most part it is removed before you get your pay check. Firms pay a number of taxes, the most obvious being *profit taxes* (or corporate taxes) on any profits made. Finally, both firms and individuals will pay taxes on any gains resulting from bank interest and profits made selling shares or property. These are collectively known as *capital gains taxes*.

2. **Indirect taxes**: Taxes that are 'hidden' or at least not as obvious are called indirect taxes, as they are paid to government indirectly. We pay for the goods and then the firms pass the tax on to government. *Import taxes* (commonly called tariffs), *specific taxes* on petrol, tobacco and alcohol – known as excise duties – and *value-added taxes* (VAT) are all examples of indirect taxation. Note that specific taxes are often set per unit, for example £2 tax per bottle of wine, whereas value-added taxes are a percentage tax based on the selling price, for example 25% on all sporting goods or 12% on foodstuffs. A percentage tax (VAT) is referred to as an *ad valorem tax*.

Figure 9.13 (WHO data) shows the total tax weight (which includes both specific taxes and VAT) on a pack of cigarettes in selected countries. The percentage of tax constitutes between 25% and 85% of the price. As smoking still seems to be commonplace in these countries, one can well imagine how government coffers are padded by these taxes. (See also "A case of stupidity study; Tobacco tax in Sweden 1997/'98" in Chapter 13.)

---

14 http://www.who.int/bulletin.; and *The Economist, Don't just sit there*, May 9th 2003; and *Obesity Tops Smoking for Medical Costs*, study compiled by Rhonda L. Rundles, Staff reporter for *The Wall Street Journal*, 1998.

15 National Audit Office of Great Britain, taken from *The "fat tax": Economic Incentives to Reduce Obesity*, Institute for Fiscal Studies; http://www.ifs.org.uk/bns/bn49.pdf

16 No, I'm not joking. The funeral tax in my country, Sweden, is 0.3% of gross income. This means that an average Swedish income earner over a life time will pay approximately SEK70,000 (circa USD10,800 in December 2008) for the Final Service. No wonder we Swedes have one of the highest suicide rates in the world; it's another way to avoid tax! (I wonder how one could measure the PED for this service.)
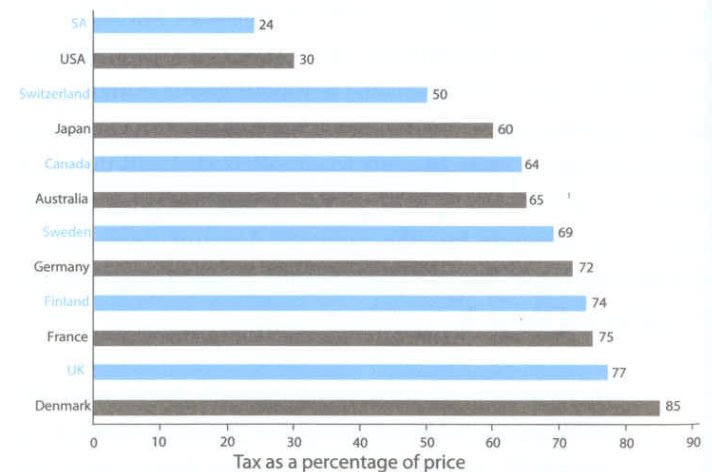

Figure 9.13 Tobacco tax as a percentage of price – selected countries

It is estimated that a 10% increase in the price of tobacco would result in a 4% decrease in quantity demanded in MDCs and around 8% in LDCs.[17] This gives a PED of between 0.4 and 0,8, (the table on page 144 gives 0.45.) which is quite price inelastic and no surprise whatsoever.

Summary & revision

1. Price elasticity of demand (PED) measures the sensitivity of demand in relation to price. PED is defined as the relative change in quantity demanded due to the relative change in price.

2. The formula for PED is: $\dfrac{\%\Delta \text{ in Qd}}{\%\Delta \text{ in price}}$.

   PED values are negative but expressed in absolute numbers, e.g. $|1|$ rather than -1.

   a. If PED > 1 the good is elastic

   b. If PED is < 1 the good is inelastic

   c. If PED is = 1 the good is unit elastic

4. Along a linear demand curve, PED values go from infinite on the P-axis to zero on the Q-axis.
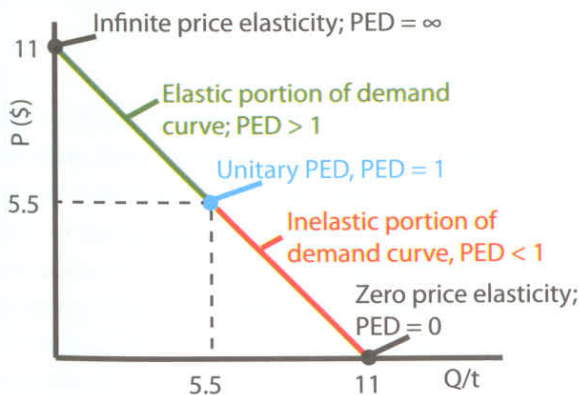
5. The determinants of PED are:

   a. Availability and closeness of substitutes

   b. Time span

   c. Proportion of income spent on the good

6. Total revenue (P×Q) along a linear demand curve is maximised when PED is unitary.
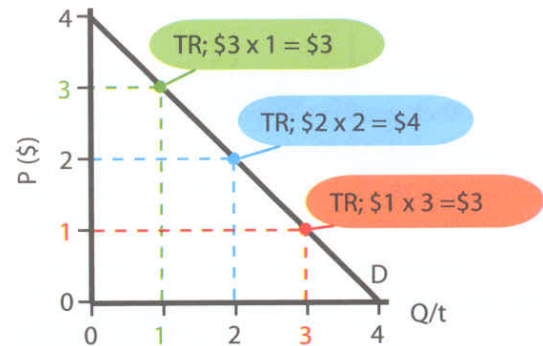
7. PED for primary goods tends to be relatively inelastic due to lack of close substitutes and/or that a small proportion of income is spent on the good.

8. PED for secondary goods tends to be relatively elastic due to greater substitutability.

9. Expenditure taxes on goods with low PED will bring higher tax revenue to government than taxes on goods with high PED.

Tax revenue – high PED

Tax revenue – low PED

MICROECONOMICS
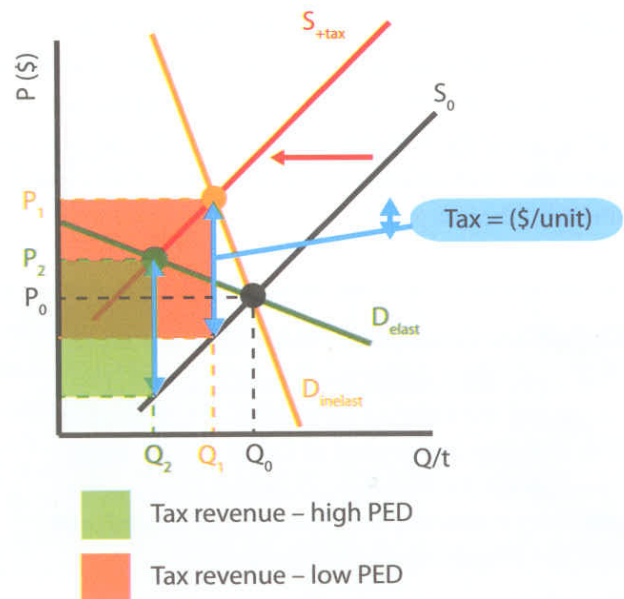
10. Cross Price Elasticity of Demand

Key concepts:

- Definition of cross price elasticity of demand
- Formula
- Significance of sign – complements and substitutes
- Applications of cross price elasticity of demand

MICROECONOMICS

Recall that many goods reveal that they have relationships with other goods. CD players and CDs are in joint demand while DVD players are in competitive demand with Blu-Ray players. This section looks at how sensitive these goods might be in terms of the price of one good affecting the demand for another. This is cross-elasticity, or as will be used here, cross-price elasticity of demand (abbreviated CPED).

Definition of cross price elasticity of demand

Cross-price elasticity is basically addressing the question 'If the price of Blu-Ray players falls, what will be the extent of change in; a) the demand for DVD players, or b) the demand for Blu-Ray discs?' Cross price elasticity is a proportional measure of the sensitivity of demand for a good with respect to a change in price of another good. In the question above, we can expect the demand for DVD players to fall because the two goods are *substitutes*. As for Blu-Ray discs, demand would increase as the two goods are *complements*.

The definition of cross-price elasticity of demand is very similar to PED. All we are doing is calculating the sensitivity of Good X with respect to *another* good, Good Y. Dividing the percentage change in quantity demanded for good X by the percentage change in the price of Good Y gives us a value of the cross-price sensitivity of the two goods.

> **Definition: 'Cross-price elasticity of demand'**
>
> The cross-price elasticity of demand – CPED – measures the relative sensitivity of a change in the quantity demanded of Good X with respect to a change in the price of Good Y. Cross-price elasticity measures the closeness of substitutes and the relevance of complements.

Formula

Assume that the price of Good Y increases by 10% which causes the quantity demanded for Good X to fall by 20%. To calculate cross-price elasticity for Good Y, we divide the percentage change in the quantity demanded of Good X (remember; 'quantity on top') with the percentage change in the price of Good Y. The *value* of cross-price elasticity will tell us whether the goods are closely related – and the *sign* whether they are complements or substitutes.

$$CPED_{x,y} = \frac{\%\Delta \text{ in Qd for X}}{\%\Delta \text{ in price of Y}} \rightarrow \frac{\frac{Q_{x_1} - Q_{x_0}}{Q_{x_0}} \times 100}{\frac{P_{Y_1} - P_{Y_0}}{P_{Y_0}} \times 100}$$

Significance of sign – complements and substitutes

Complements

Complement goods are goods where the consumption of one good leads to the consumption of another good, such as the classic 'tennis balls and tennis racquets', which is a singularly uninspiring example. I will instead vent my spleen (= express my feelings strongly) using a particular pair of products that has long yanked my chain (= irritated me), namely printers and printer ink cartridges.

When looking at printers, one will of course look at the total cost picture; a printer needs ink cartridges. Say that the price of ink cartridges fell from €15 to €13.5 (-10%) and that subsequently demand for printers increased from 10,000 units per week to 12,000 (+20%). The cross-price elasticity (X = printers; Y = ink cartridges) of printers with regard to cartridges would be:

$$\frac{\dfrac{Q_{x_1} - Q_{x_0}}{Q_{x_0}} \times 100}{\dfrac{P_{Y_1} - P_{Y_0}}{P_{Y_0}} \times 100} \rightarrow \frac{\dfrac{12000 - 10000}{10000} \times 100}{\dfrac{135 - 150}{150} \times 100} \rightarrow \frac{+20\%}{-10\%} = -2$$

Notice the appropriate signs in the formula and answer. Take heed; it is most important to assign these signs correctly to cross price elasticity values. The *minus sign* tells us that we are dealing with **complement goods**; as the goods are complementary a decrease in the price of one good will cause an increase in the quantity demanded of the other. Any positive value divided by a negative value – or vice versa – will render a negative value. The value of minus 2 also tells us that printers and ink cartridges are highly complementary, known as *strong complements*. The higher the value of CPED (in absolute terms, i.e. disregarding the minus sign) the stronger is the joint demand. Note that even if the price of ink cartridges were to rise and the quantity demanded for printers were to fall, we would still get a negative value.
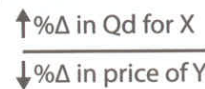
Substitutes

Let us continue with the example of DVD players and Blu-Ray players. Assume that the price of Blu-Ray players falls. This would most certainly have an effect on the demand for DVD players as the two goods would seem to be highly substitutable. Say that the price of Blu-Ray players went from USD200 to USD180 and that the quantity demanded of DVD players decreased from 10,000 units to 9,500 units during a given time period. Our cross-price elasticity (X = DVD players; Y = Blu-Ray players) would be:

$$\frac{\dfrac{Q_{x_1} - Q_{x_0}}{Q_{x_0}} \times 100}{\dfrac{P_{Y_1} - P_{Y_0}}{P_{Y_0}} \times 100} \rightarrow \frac{\dfrac{9500 - 10000}{10000} \times 100}{\dfrac{180 - 200}{200} \times 100} \rightarrow \frac{-5\%}{-10\%} = +0.5$$

A **positive value** of 0.5 tells us that the goods are **substitutes** – but a value of less than one tells us that they are relatively weak substitutes. A lower price of Blu-Ray players led to a decrease in quantity demanded for the substitute, DVD players. As no surprise, if the price of one had risen, the quantity demanded of the other would have increased which means that the cross-price elasticity would still be positive.
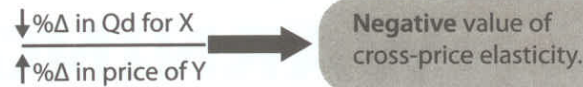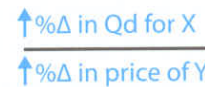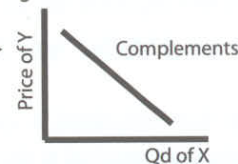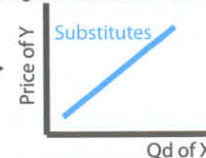

Figure 10.1 Summary of cross-price elasticities

Accept no substitutes! (On the road in the Yucatan, Mexico, July 2007)

Diagrammatical analysis of substitutes and complements

You must be able to use diagrams in economic analysis and cross-price elasticity is no exception. Here are two examples of how we could use supply and demand analysis to show the relationship between complementary goods and substitute goods.

DANGER MINES

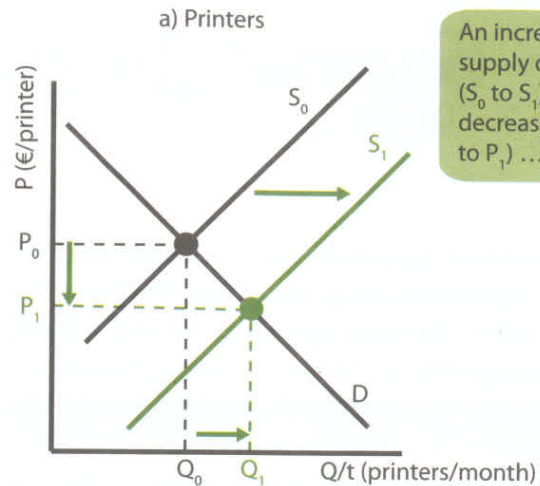# WARNING !

CPED and 'ΔQd' vs. 'ΔD'

One of my economics colleagues at the Oxford Study Courses took issue (in a very polite British manner) with my illustrations of the correlation between the change in price of Y and the change in quantity demanded of X shown in Figure 10.1. Stephen is a super, modest, unpretentious bloke (and commands great respect from his students who invariably call him "Mr Holroyd") so I shan't name him.

Stephen pointed out that using a diagram with 'Price of Y' on the Y-axis and 'Quantity demanded of X' on the X-axis would be '… invariably confusing and borderline erroneous…' since a change in price of a good for which there is a complement or substitute good will shift demand for said complement/substitute. Yes, this is correct. My defence in using the correlative diagrams ($P_Y \Leftrightarrow Qd_X$) is simply to illustrate what the formula is based on; % $\Delta P_Y \rightarrow$ % ΔQd_X.

I recommend that you do as Stephen says; don't use the explanatory diagrams in Figure 10.1 but the ones further on (Figures 10.2 and 10.3). I want no truck (= US slang for 'trouble') with a 195 cm man who outweighs me by some 30 kg and spends his free time hauling 10 kg honey pots – the man has muscles in his teeth!

MICROECONOMICS

a) Printers

An increase in supply of printers (S_0 to S_1) leads to a decrease in price (P_0 to P_1) …
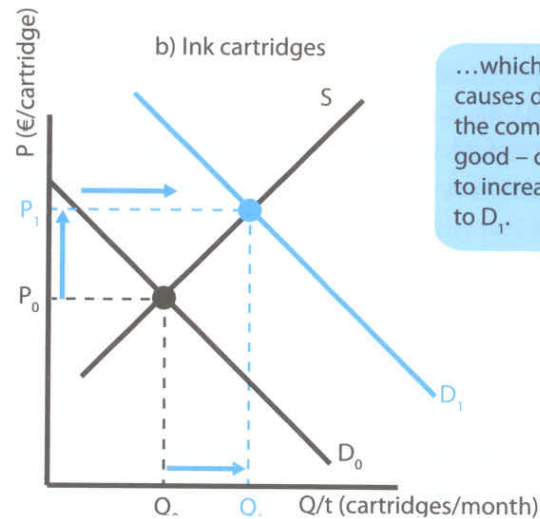
b) Ink cartridges

…which in turn causes demand for the complement good – cartridges – to increase from D_0 to D_1.

Figure 10.2a) & b) Complement goods – printers and ink cartridges

Complement goods – shifting demand

I continue with the two complementary goods, printers and ink cartridges. In Figure 10.2a the supply of printers increases, say due to increased productivity at the firms producing the players. This lowers the market price of printers and increases demand for the ink cartridges used in the printers, Figure b.

Substitute goods – shifting demand

As for substitute goods, DVD players and Blu-Ray players are of course in competitive demand. In Figure 10.3a and b, the supply of Blu-Ray players has increased, say due to lower production costs in manufacturing. This lowers the market price of Blu-Ray players and decreases demand for DVD players which are substitutes.

a) Blu-Ray players

P (€/Blu-Ray player)

An increase in supply of Blu-Ray players (S0 to S1) leads to a decrease in price (P$_0$ to P$_1$) …

b) DVD players

P (€/DVD player)

…which in turn causes demand for substitutes – DVD players – to decrease from D$_0$ to D$_1$.

Figure 10.3a) & b) Substitute goods – Blu-Ray players and DVD players
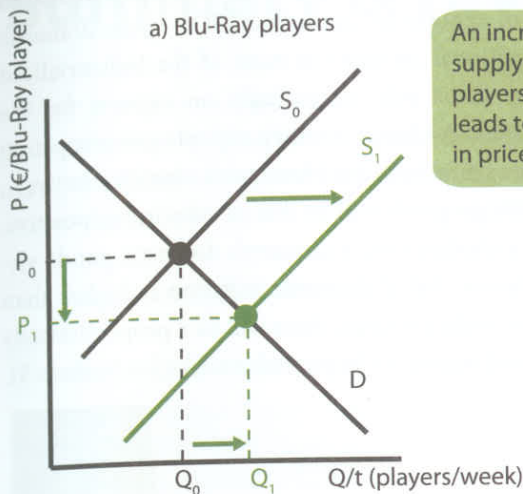
OUTSIDE THE BOX

Recall how the demand for a good is influenced by the price, availability and demand for other goods; substitutes and complements. A substitute is a good which is in competitive demand with another, while a complement good is in dual demand with another good. Let us go through the possible substitutes and complements.

Many of my students are avid fans of both Harry Potter and the Lord of The Rings trilogy (which is in four parts, for some reason). We have had many a discussion on the subject of the relationship between books and films based on books. We have arrived at some sort of consensus, as we have agreed that these two goods are in fact both substitutes and complements. Our argument goes something like this: if one has read a book and really enjoyed it then chances are one would want to see the film in order to see whether the film is true to the book. These two goods would thus be in joint demand – they are in fact complement goods.
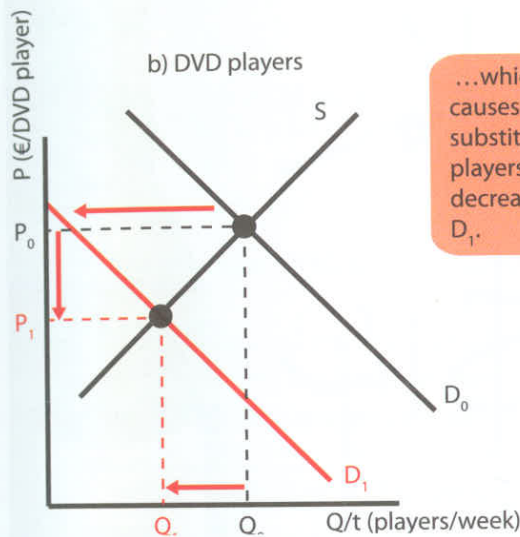
Having said this, one can also make a strong case for saying that they are substitute goods! How often have we seen a film but not read the book? And having seen the film, do we then read the book? Good question! There is an impulse to think 'Naaah. Been there, done that, got the t-shirt.' Yet there is also an incentive to find out if the book is better – as is so often the case – so one might want to get hold of a copy. Basically, the question is whether a film made from a book will complement the book by creating demand for it, or whether the film can act as a substitute for the book. A most un-scientific inquiry amongst my students paints a picture of a stronger complement effect that substitution effect – which incidentally might help to explain why so many recent editions of both the Harry Potter books and The Lord of the Rings trilogy have the film characters on the cover!

A little rumination… Harry Potter and the Economist's Stone

The much-hyped return of the young necromancer Harry Potter is an excellent example of the relationship between different goods offered on a market. So good in fact that this is the third time I am using it in versions of this book! The J. K. Rowling series of books has been a fabulous success and the first film released in 2001 was a smash hit – and the seventh book has been made into two films to squeeze the very last money drop from the series. Let us have a look at some of the market issues.

POP QUIZ 10.1

1. Explain how a change in the price of good X can cause both price and quantity to increase for complement good Y.

2. A certain product has very few substitutes and a very small proportion of consumer income is spent on the product. What would the PED be like?

93

3.   The price of a municipal bus fare rises from 80 pence to 90 pence and quantity demanded falls by 10%. What is the value of price elasticity of demand?

4.   Assume that apples and oranges have a positive cross elasticity of demand whilst oranges and orange peelers have a negative cross elasticity. How would you define the two separate 'pairs' using economic terminology?

5.   Letters and email have a cross price elasticity of 3. If the price of postage rises from 4 DKK (Danish Crowns) to 6 DKK, how much will demand for email change?
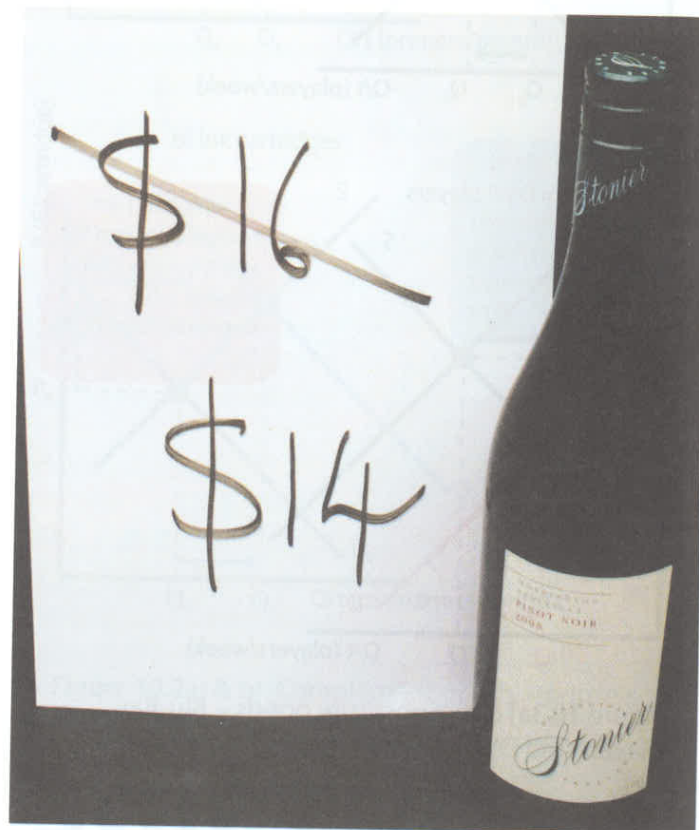
Applications of cross price elasticity of demand

The sensitivity of quantity demanded in relation to a change in the price of related goods will have a major impact on pricing and output decisions of firms. The example of DVD players and DVDs used earlier illustrates how suppliers of discs will have to be well aware of changes in the market price of players in order to compensate output and/or prices of discs. Knowledge of the cross-price elasticity of these *complement goods* will enable the disc firm to plan ahead.

The same is true for *substitute goods*. If a competitive good, say Blu-Ray players, falls in price then suppliers of both DVD players and Blu-Ray discs will need to know how their own good will be affected and by how much. Again, producers will want to plan for eventual production changes by increasing or decreasing factor inputs, e.g. labour and capital. A main element in knowing the effect on demand is being able to budget the operation correctly. Both revenue and profit will be affected by any change in demand, so firms put a great deal of effort into correctly estimating future demand. Many firms will in fact have entire divisions whose purpose is to analyse potential opportunities and threats to their product. Threats might be the new revolutionary product of a rival firm which would threaten own sales. Opportunities might be being able to purchase the patent rights, production rights, licence rights or entire rival company. This is the case very often in the highly competitive market for smart phones (which are more like hand-held computers), where rivals often buy the license to use software or simply merge.

Cross-price elasticity of demand also helps to understand the increasing importance of the ongoing process of *international*

integration. Improved ease of cross-border trade has increased the proportion of exports far faster than GDP (gross domestic product – e.g. national income) in most of the industrialised countries. Many firms rely increasingly on exports for the brunt of their sales. As home country exports are competing with many other countries, the cross-price elasticity between domestic and foreign goods will be that of substitutes; positive. High cross-price elasticity for a country's domestic goods vis-à-vis imports means that if domestic inflation is higher than foreign countries' inflation rates, there will be a proportionately larger increase in demand for imported goods. (See Section 3).

Now what would go well with this? How much cheese can one get for $2?

Summary & revision

1. Cross price elasticity of demand (CPED) shows the sensitivity of demand for a good with respect to the price of

   a. a complement

   b. a substitute

2. The formula for CPED is

$$CPED_{x,y} = \frac{\%\Delta \text{ in Qd for X}}{\%\Delta \text{ in price of Y}} \rightarrow \frac{\dfrac{Q_{x_1} - Q_{x_0}}{Q_{x_0}} \times 100}{\dfrac{P_{Y_1} - P_{Y_0}}{P_{Y_0}} \times 100}$$

3. A positive value of CPED means the goods are substitutes.

   • Substitutes in diagrams; a decrease in the price of Blu-Ray players will lead to a decrease (demand curve shifts left) in demand for DVD players.

4. A negative value of CPED means the goods are complements.

   • Complements in diagrams; a decrease in the price of DVDs will lead to an increase in demand (demand curve shifts right) for DVD players.

11. Income Elasticity of Demand

Key concepts:

- Definition of income elasticity of demand (yED)
- Formula for income elasticity of demand
- Normal goods
- Inferior goods
- Applications of income elasticity of demand

The concepts income and demand are inextricably linked, just like having a good day on the stock markets and celebrating with a nice cigar and a bottle of Champagne. One is unlikely to celebrate by going out and splurging on, say... potatoes![1] This is what income elasticity looks at; the propensity to change one's consumption of certain goods due to a change in income. Income elasticity of demand measures how responsive quantity demanded is to a change in income.

Definition of income elasticity of demand (yED)

The issue at hand is to relate a change in quantity demanded with an increase in income, i.e. the *percentage change in quantity demanded over the percentage change in income*. Most goods will behave normally, which is, the quantity demanded will increase as income increases. However, some goods will actually behave in an opposite manner; quantity demanded will fall when income rises. Therefore, just as in cross-price elasticity, we are interested not only in the *value* of elasticity, but also the *sign*.

> **Definition: 'Income elasticity of demand'**
>
> Income elasticity of demand – yED – measures the responsiveness of demand, i.e. the relative (percentage) change in the quantity demanded for a good due to a relative (percentage) change in income. (The lower case 'y' signifies personal income, since upper case 'Y' is reserved for national income.)

Formula for income elasticity of demand

The formula for income elasticity is nothing new – we simply replace 'relative change in price' with 'relative change in income'. I urge you again; beware of the sign of yED!

$$yED = \frac{\%\Delta \text{ in Qd}}{\%\Delta \text{ in y}} \rightarrow \frac{\frac{Q_1 - Q_0}{Q_0} \times 100}{\frac{y_1 - y_0}{y_0} \times 100}$$

Normal goods

The last time I almost resigned (or was almost fired – these seem to be complement goods in my case) I stormed home and told my then wife that a number of planned expenditures would have to be dropped. The mere whiff of less income caused a change in purchasing behaviour. As soon as I was re-hired (with the mediator help of Joe Cool, the Irish math teacher) my household's buying plans went right back on track. This is how consumers are prone to act – and all the goods whose quantity demanded fell and rose in concert with my income are normal goods. The definition of a **normal good** is that it is positively related to income. Most goods are normal goods, but there are a few possible exceptions.[2]

> **Definition: 'Normal good'**
>
> A normal good is positively correlated to income, i.e yED is positive. A rise in income increases demand for normal goods. A fall in income decreases demand for normal goods.

---

1   Unless you're Swedish or Irish. I'm a bit of both and I can tell you; we accept no substitute.

2   This goes for school headmasters also.

An *income-elastic* good is a good for which quantity demanded rises proportionately more than an increase in income, i.e. income elasticity is greater than +1. High market goods such as vintage wines and luxury goods are income elastic, as are services and tourism. At the other end of the spectrum are goods where the increase in quantity demanded is proportionately less than the increase in income; *income inelastic* goods, which have values of income elasticity less than +1. Commonly referred-to examples are clothing and cigarettes.

I figure you are all big boys and girls now, so I shall forgo all the formulaic usage and confine myself to examples based on percentages. Let us assume that income for a specified group of consumers rises by 10% - which is a bit unrealistic but makes the calculations easier. If quantity demanded for vintage wines from Bordeaux increases by 20% we get a yED of +2; an increase of vegetable consumption by 2% renders a yED of +0.1.

Quite naturally, income elasticities will vary between high and low income groups and more significantly perhaps, between high and low income countries. Food, for example, has an income elasticity of between +0.5 and +0.6 in LDCs while most MDCs hover between +0.2 and +0.5.[3] What might be considered a luxury good in rural China, a bicycle for example, would have a correspondingly high income elasticity of demand whereas a MDC would most assuredly have a lower income elasticity of demand for such a good. In line with this, development and accompanying higher incomes will change the pattern of income elasticities away from basic commodities towards manufactured goods and services.

Inferior goods

A good where quantity demanded falls as income rises is known as an **inferior good**, the term being used not as a descriptor of the good per se, but rather as a descriptor of our relative consumption preferences at higher income levels. Potatoes are not 'inferior' to pasta yet when incomes rise evidence suggests that our relative preferences will cause us to actually lower our consumption of potatoes and substitute the good with something preferable. Potatoes might thus have a negative income elasticity of demand, as an increase in income is correlated to decreased quantity demanded.

> **Definition: 'Inferior good'**
> An inferior good is negatively correlated to income, i.e. the yED is negative. A rise in income decreases demand for inferior goods. A fall in income increases demand for inferior goods.

It is worth commenting on the possibility that many goods are in fact normal up to certain incomes and thereafter inferior. At low levels of income, any addition to income might mean increased quantity demanded for basic goods such as food and clothing. However, at some point at higher level income, the quantity demanded for such goods would surely not proportionally match an increase in income. Just imagine how the yED values for economics students at university would change upon getting a good job in London City after graduation.

An oft-cited concept within the context of inferior goods is the aforementioned Giffen good, i.e. a good where quantity demanded is *positively* correlated to price. This would be a type of inferior good… but I've just decided that it would be a waste of time explaining this, as Giffen goods invariably wind up in the same category as mermaids, goblins and UFOs. In the previous edition of this book I declared an international boycott against the use of the term 'Giffen Good'. Somebody in the IBO seems to have caught on since this ridiculous concept has since disappeared from the syllabus! I rule.

POP QUIZ 11.1

1. How will demand for inferior goods in an economy be affected if a) incomes rise, b) unemployment decreases c) the general price level rises? (Treat each as a separate case.)

2. You read in the newspaper about a good with '…an elasticity of -2'. Explain why you need more information about the type of elasticity since '-2' can actually be referring to two entirely different goods!

3. Must income elasticity of demand be solely positive or solely negative? Illustrate your answer with a diagram.

---

3 International Evidence on Consumption Patterns 1989, Greenwich, Connecticut JAI Press Inc.

Applications of income elasticity of demand

Summary & revision

Income elasticities play an important roll in illuminating the problems of less developed countries (LDCs) caught in producing primary goods. A solid grasp of how income affects consumption preferences is central to understanding how economic growth influences trade and development.

LDCs are highly dependent on the production of primary commodities. 71% of non-Asian LDCs' exports are made up of primary goods.[4] Yet the same countries account for just 12% of total world trade and 17% of total world GDP (1998 figures).[5] Summing this up paints a picture of a group of countries whose output is exported mainly to rich countries, and where the demand for these primary goods is highly dependent on growth rates in high-income countries.

- As average global incomes increase, demand for goods with high yED (secondary and tertiary goods) will rise faster than for goods with low yED (primary goods).

- This paints an important picture of *relative* prices; the price of secondary goods will rise faster than prices of primary goods. In economic lingo, primary goods become *relatively* cheaper than secondary goods.

- Economies specialising in primary goods will now see that increasing quantities of, say, coffee must be exported in order to import any given amount of machinery. The price of exports relative to the price of imports is known as the *terms of trade*. As export prices fall in relation to import prices, the terms of trade worsen for primary goods specialists.

1. Income elasticity of demand (yED) shows the sensitivity of demand for a good with respect to a change in income

2. The formula for yED is

$$yED = \frac{\%\Delta \text{ in Qd}}{\%\Delta \text{ in y}} \rightarrow \frac{\frac{Q_1 - Q_0}{Q_0} \times 100}{\frac{y_1 - y_0}{y_0} \times 100}$$

3. A positive value of yED means the good is a normal good

   Normal goods in diagrams; an increase in income leads to an increase in demand (demand curve shifts right)

4. A negative value of yED means the good is an inferior good

   Inferior goods in diagrams; an increase in income leads to a decrease in demand (demand curve shifts left)

5. LDCs are often producers of primary goods which have low income elasticities of demand. As incomes in developed countries increase, demand for primary goods increases proportionally less. As prices for secondary goods – which have a higher income elasticity of demand – rise more than prices for primary goods, LDCs will have to export more primary goods for any given quantity of imported of secondary goods. This means that the terms of trade for LDCs worsens.

---

4   Todaro, p 68

5   *The World Economy, A millennial perspective*, OECD, table 3-1c, page 127

12. Price Elasticity of Supply

Key concepts:

- Definition of price elasticity of supply
- Formula
- Range of values
- Diagrammatical illustration
- Determinants of price elasticity of supply
- Applications of price elasticity of supply

This chapter addresses the issue of suppliers' *ability* and *willingness* to put a good on the market. The price elasticity of supply looks at the responsiveness of supply to a change in demand – e.g. how quickly firms can adjust supply to meet a change in demand. The measurement of this is the **price elasticity of supply**.

Definition of price elasticity of supply

Compare the three goods in Figure 12.1. Why would a supply curve be 'steep' or 'shallow'? Try to fit the correct good to each diagram; blank DVDs, the Olympic boxing finals and the supply of IB economics hours in a public school (answer in footnote).[1] It is rather self-evident that the supply curves indicate the willingness/ability of firms to increase output of a good within a given period of time.

---

1. The answers are: Figure a = blank DVDs, Figure b = supply of IB economics, Figure c = supply of tickets to the Olympic boxing finals. You may still send me $5.
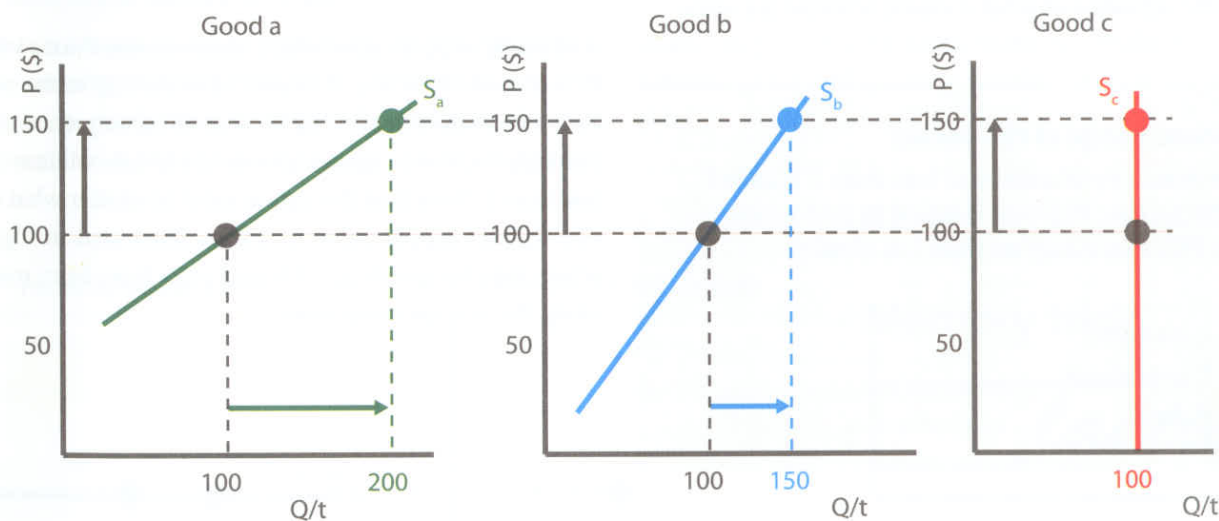

Figure 12.1 Supply curves for three goods

The diagram series above pretty much says it all. The price has increased by 50% in all three cases (from $100 to $150) but the percentage increase in quantity supplied varies from 100% to zero. We have just defined **price elasticity of supply** (PES); it is the *relative increase in quantity supplied* due to a *relative increase in price*. The price elasticity of supply (PES) for **Good a** is 100/50; PES = 2: **Good b** is 50/50; PES = 1.0. **Good c** is 0/50; PES = 0.

MICROECONOMICS

Definition: 'Price elasticity of supply (PES)'

The price elasticity of supply (PES) is a measure of the responsiveness of firms in increasing the quantity supplied on the market due to a change in price. The supply curve is upward sloping, i.e. PES will be positively correlated to price, with few exceptions.

Formula

By now you should recognise the formula methodology. The only difference between the formula for PES and PED is that we replace Qd (quantity demanded) with Qs (quantity supplied).

$$PES = \frac{\%\Delta \text{ in Qs}}{\%\Delta \text{ in price}} \rightarrow \frac{\frac{Q_1 - Q_0}{Q_0} \times 100}{\frac{P_1 - P_0}{P_0} \times 100}$$

Range of values

Just as demand elasticity will have an endless range, so will supply elasticities. The main difference is that elasticity of supply will not have a range of infinity to zero along a single curve. Yet the *definition* of what is elastic remains the same.

Definition: 'Range of PES values'

A price elasticity of supply of less than 1 is called inelastic supply; PES that is equal to one is unit elastic; PES that is higher than 1 is elastic.

PES < 1; inelastic

PES = 1; unit elastic
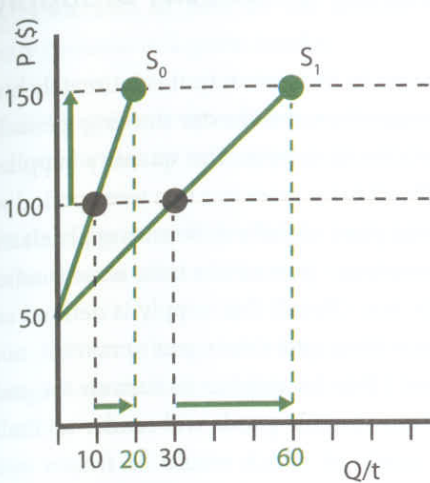
PES > 1; elastic

Diagrammatical illustration

Supply curves S_a, S_b and S_c in Figure 12.2 show the differences in price elasticities of supply for the three goods. As in PED, the values of PES will vary along the curves, yet depending on whether the curve intercepts the P-axis, Q-axis, or origin, there is a boundary to the values.
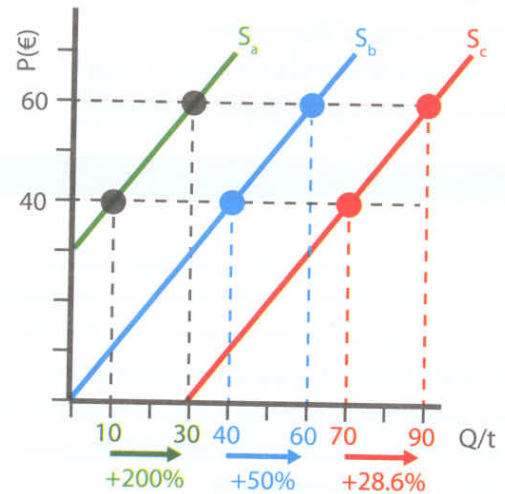
PES$_a$: 200%/50% = 4

PES$_b$: 50%/50% = 1

PES$_c$: 28.6%/50% = 0.57

Figure 12.2 PES for three goods

Any linear supply curve which starts on the P-axis will have a PES of **more** than one. A linear curve starting at the origin will have a PES **equal** to 1 and in full logical consequence; a straight-line supply curve originating from the Q-axis will have a PES of **less** than 1. Note that this rule is valid no matter what the slope of a (linear) supply curve is. This is illustrated in Figure 12.3, where the changes in quantity and price have been marked out along the axis for comparison.

PES > 1; supply is elastic
%ΔP = 50;
%ΔQS = 100

A straight-line supply curve originating from the P-axis will always have a PES > 1, since the percentage change in quantity supplied will always be larger than the percentage change in price.
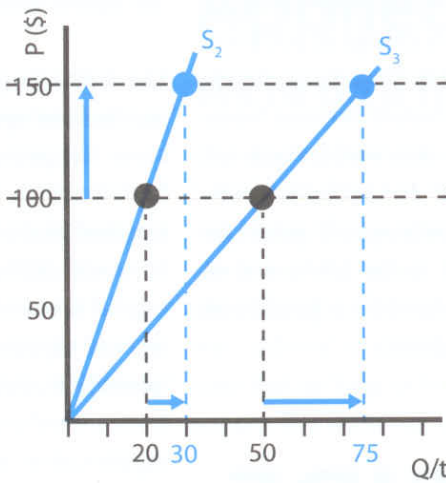
PES = 1; supply is unitary
%ΔP = 50;
%ΔQS = 50

A straight-line supply curve originating from the origin will always have a PES = 1, since the percentage change in quantity supplied will always be the same as the percentage change in price.
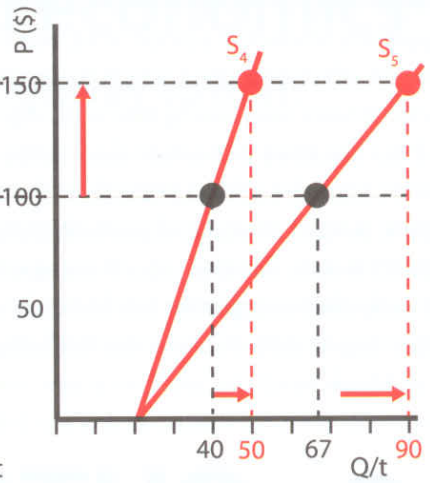
PES < 1; supply is inelastic
%ΔP = 50;
%ΔQS = 25

A straight-line supply curve originating from the P-axis will always have a PES < 1, since the percentage change in quantity supplied will always be less than the percentage change in price.

Figure 12.3 PES for different axis-intercepts

A few extreme cases of PES

There are two extreme cases of price elasticity of supply. One is where the quantity supplied remains the same no matter what the price, i.e. *perfectly inelastic* supply. The other case is when quantity supplied is infinite at one specific price, i.e. *perfectly* elastic supply. In both cases, there is actually no correlation between price and quantity supplied. Figure 12. 4 illustrates this.
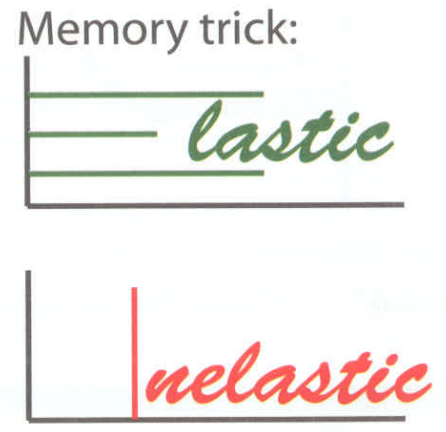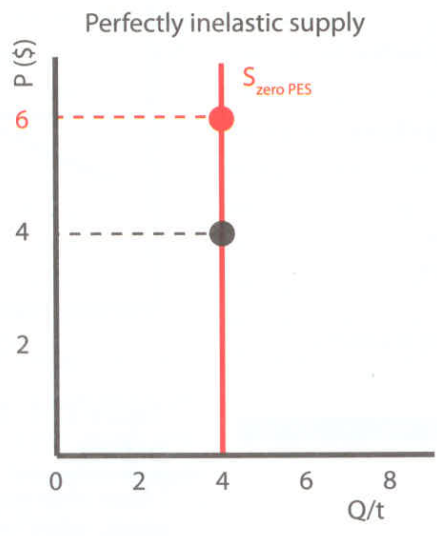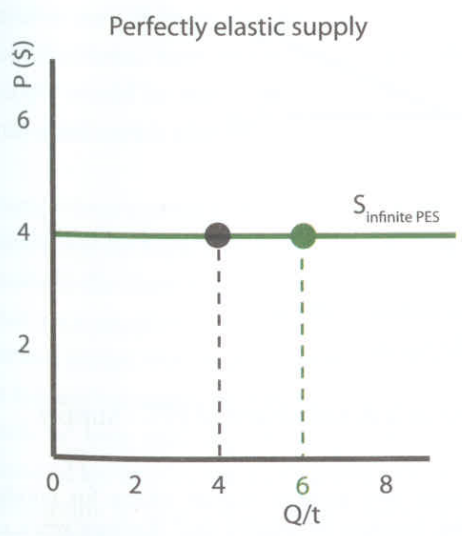

Figure 12.4 Extreme cases of PES

When supply is *perfectly elastic*, suppliers are able/willing to supply any amount at a price of $4, while a change in price will have no effect on a good which has a *perfectly inelastic* supply curve. It bears mentioning that while equivalent extreme forms of PED are most difficult to find in the real world, there are several examples of both perfectly inelastic and perfectly elastic supply. A valid example of perfectly **in**elastic supply would be concert tickets (Chapter 15, 'Born again in the USA') and of perfectly elastic supply the world supply curve for a good in an importing country (Chapter 65, tariff diagram).

OUTSIDE THE BOX

I have taken care to use the term 'straight-line' or 'linear' supply curve in explaining how the value of PES depending on where the curve originates. The reason is illustrated in Figure 12.4, where a curve with an increasing slope shows that point values of PES along the curve change. Following section S1 through the tangential points S_2, S_3, S_4 to section S_5, PES goes from infinity to zero. This shape of the supply-curve has relevance for the total output curve (= aggregate supply curve) in Section 2, macro.
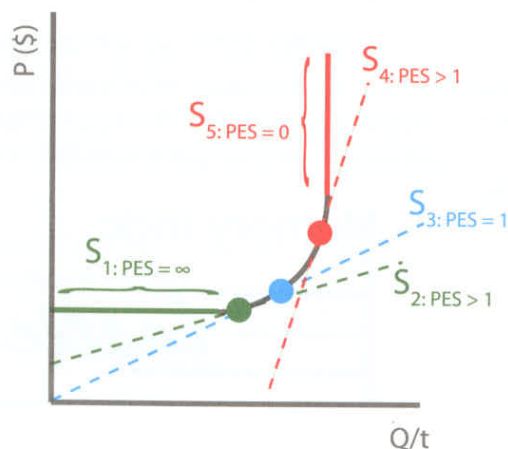
Determinants of price elasticity of supply

The ability for producers to increase output is intimately hinged on the **time span** in question; the shorter the time period, the lower the price elasticity of supply. The quantity supplied of fresh tomatoes is difficult to increase in a two week period while a period of six months gives entirely different supply elasticity. So would canned tomatoes – part of the time issue is whether a good is **storable** or not. (Recall that supply is defined as the amount suppliers are willing and able to put to market, not the 'quantity in existence'.) It is far quicker to furnish the market from a warehouse. Non-storable goods will render an inelastic supply-curve, an example of which would be frozen oysters. Fresh oysters are an excellent example of a good which cannot be stored for any length of time.[2]

This makes sense when one thinks about how the difficulties in increasing the supply would be alleviated over time. Increasing the quantity of beef supplied per month is far more difficult that increasing the supply over a two year period. The producers will have time to come up with better cross-breeding methods and expand herds. All this would increase producers'/suppliers' ability to increase the quantity supplied within the time period. Figure 12.5 shows how supply would differ over time; immediate supply ($S_{immediate}$) would be perfectly inelastic, supply in the short run (S_{SR}) would be inelastic, supply in the long run (S_{LR}) would be elastic and supply in the very long run (S_{VLR}) completely elastic. The diagram also serves as a reminder of how the short run supply inelasticity of *primary goods* such as agricultural produce, iron ore, and oil causes large price fluctuations.

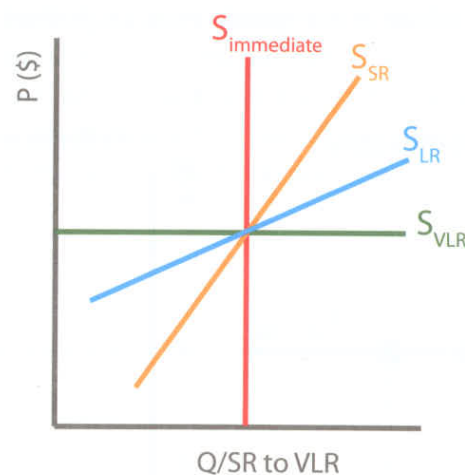Figure 12.4 Changing PES values along a supply curve

Figure 12.5 Time as a determinant of PES - supply of beef

2   I have observed some *very* strange market effects for goods which, when stored, *increase in quality* and therfore price – such as Cuban cigars and Rioja wines. Somebody needs to do an extended essay on this.

$S_{immediate}$: The PES of beef in the immediate term would be perfectly inelastic in a given market.

S_{SR}: Over longer time spans however, increases in demand can more easily be met by increased stocks of frozen meat.

S_{LR}: In the long run, larger herds and increased land use for cattle (another re-allocation issue!) increases the PES.

S_{VLR}: In the very long run, improved genetic cattle strains resulting from crossbreeding and genetic modification increase PES further.

In a similar vein, time is an issue for producers who are at the limits of **output capacity**. The availability of excess capacity, say available machines, labour and factory space, will have a major impact on the ability to increase supply within a given time frame. In the short run it can be difficult to get hold of scarce factors and to expand the size of production plants. Over longer time-periods it is easier to plan and target output by increasing bulk-buying of material and increasing the amount of capital used.

Closely related to the issue of the time involved in increasing quantity supplied within a given period, is the ability (and cost!) incurred by producers in switching from the production of one good to another. A *producer substitute* is a good which suppliers can switch to as an alternative – not to be confused with consumer substitutes. The ease or difficulty any given supplier experiences in moving productive resources from one good to another will be a major factor in determining supply elasticity. A common classroom example is whiteboard pens and permanent markers being very close producer substitutes. Switching from whiteboard pens to permanent markers[3] would involve very little in the way of tool readjustment, machine modifications, new material and knowledge and so forth. Supply would be very elastic. Switching from compact cars to mid-sized cars is also fairly supply elastic.

Many *primary goods* such as iron ore, teak wood and agricultural goods will be highly inelastic, as the ability to switch to such goods in the short run is very limited. The basic premise here is that once suppliers have a large amount of resources sunk in a certain sector, the re-applicability of these resources is the key to increasing supply of other goods. Just consider commodities such as iron and copper; if you have a copper mine and demand for iron increases it's going to be a bit difficult to switch production to iron. I mean, you have a *copper* mine…

3 A favourite classroom prank known to all my people, incidentally.

Applied economics
We're on a highway to hell!

One day Joe the math teacher came down to my office to ask me 'what the heck I was doing' in my IB1 class 'and would I please stop doing it'. Apparently, I had made the claim that we would never run out of oil. Ever. Joe wanted to take issue with this statement. I explained that I was not making any claim that oil was 'infinite' as a resource but that market forces would make sure that as long as there was a need for oil, there would be a supply. The stone-age did not end because we ran out of stones.

What happened in November of 1973, was that OPEC – for political reasons – decreased supply. As demand was highly inelastic in the short run the price of oil went from USD3 per barrel to short run equilibrium of USD11 during 1974, a 260% increase in price. The next 'oil crisis' in 1979/'80 was caused primarily by the ripple-effects of negative expectations as the Iran-Iraq war broke out. The price rose even further, but not the shock level of '73/'74 as the PED of oil had increased.

We have been running out of oil since my grandfather drove buses in St. Louis during the depression. We have been running out of oil since the 2nd World War; since the 1960s; since the first and second oil crises in '73/74 and '79/80…you get the picture. In 1940 there was an estimated 10 years of consumption left, in 1980 over 25 years, and in 2000 40 years! The real price of a barrel of oil in 2000 was the same as in 1920. The Bronze Age did not end because we ran out of copper and tin.

This is a classic example of the 'horse-shit theorem'. I read somewhere that had computers existed in the mid 19th century, they would have predicted that by the end of the 21st century the world would be covered in six feet of horse-shit. This is the danger of extrapolation; basing a prediction on past and current facts/happenings. What has happened, of course, is that we have moved on to other forms of transportation. Just like stones, horses, and the original Donkey Kong games, we have moved on. Yes, we are running out of oil. We would run out of oil ultimately even if we only used 10 litres per year – that is the definition of 'finite resource'. **Yet the fact remains: Oil supply is not a question of physics but of economics.** The Iron Age did not end for a lack of iron.

MICROECONOMICS

103

We are not running out of oil for three reasons:

1. The price and incentives function of supply and demand has made sure that when oil prices have skyrocketed, there has been a major incentive to find new fields and to use the old fields more efficiently by developing technology. The simple fact is that previously costly oil resources became economically viable when the oil price rose. Enormous fields were found in the 1980s and 1990s; As price increases, there is an incentive to find new fields and get oil out of previously non-economically viable well/fields. These include the North Sea, Russia, Indonesia, Mexico, Canada, Alaska, to name a few.

   The high price of oil has also led to previously unprofitable shale oil deposits (oil locked in sand) becoming increasingly interesting. Estimates of the amount of oil held in shale deposits in Canada alone could increase total known reserves by 50% - the problem being that it is more expensive to extract, about 30 USD per barrel rather than 8 – 15 in Alaska and Saudi Arabia. If the price of oil increases enough, then it becomes economically viable to extract the shale oil. Estimates of total reserves of shale oil would take care of the world's TOTAL energy needs (at present rate of consumption) for the next 5,000 years.

2. 'Oil crisis' never meant that there wasn't enough oil, but that supply had fallen. Scarcity does not mean 'not enough oil' but 'dear oil'. As oil became expensive during the '70s and '80s oil crises, the rationing function and substitution effects kicked in. The rationing effect meant that quantity demanded fell and the substitution effect meant that people started to find substitutes and ways to become less dependent on oil.

3. The reasons outlined above have led to far greater efficiency in both the extraction and use of oil. We are simply getting better in production and use. In 1973, 38% of the world's oil came from the Middle East – now it's below 30%, in spite of a 25% increase in demand. Technology in oil production means that when previously 20% of the oil in a field could be withdrawn we can now withdraw over 30%. Estimates show that America's 10 largest oil fields held over 60% of their oil when the fields were closed down. These 'known reserves' increase by the year – the increase in known reserves is about 900% since 1950 and by close to 50% since the mid '70s. The cost of extracting the aforementioned shale oil in Canada has dropped from USD30 a barrel in the '70s to less than USD12 a barrel in 2002. On the users side; cars have increased efficiency by over 50% since mid-1970s, houses are better insulated and we have become much better at recycling plastics and other materials which use oil in their production.

In summary; the supply of oil has been increasing monumentally while our demand, although increasing, has become more elastic as we have started to seriously look at substitutes, such as natural gas and also renewable resources such as geo-thermal energy, wind and solar power etc. Again, we started to look for substitutes as soon as renewable energy options became competitive with oil. 35 years after the first oil crisis, the price of oil was marginally higher in 2000 than it was in 1973. The present Information Age will not end because we run out of information.

Oil is not going to become scarce. It is going to become uninteresting.

During the 1980s and 1990s, demand for oil continued to rise and supply was able to not only keep up but outstrip demand increases to the extent that the price of a barrel of oil dropped to the historic low of USD10 a barrel during 1997/98, when the real price of oil was the lowest ever.

Addendum to this edition: During the writing of this edition, a couple of colleagues were again sniggering over the fact that this 'Applied economics' box would have to be scrapped – or severely re-done. The underlying reasoning was a mixture of '…we've passed 'peak oil production'…oil prices are at a permanently high level…"…etc. I continued to argue that there will *always* be enough oil – or in fact, *too* much oil, given environmental issues – and that this 'Applied economics' box would be kept in all coming editions until I was proven wrong. In February 2011, I looked at the diagram below. You tell me; is it time to scrap this box?
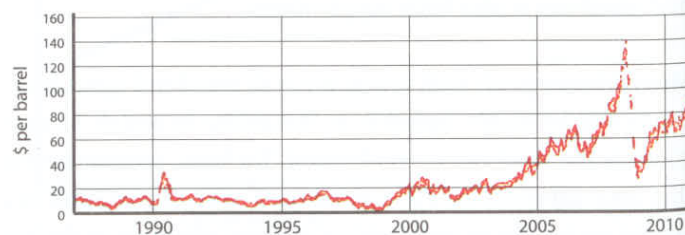

Figure 12.6 Daily European Brent Spot Price FOB
Source: http://tonto.eia.doe.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EER_EPD2DC_PF4_Y05LA_DPG&f=D

(Sources: Most of the figures and statistical data have been taken from a truly astounding book, *The Skeptical Environmentalist*, Björn Lomborg, pages 118 – 138. Additional figures on shale deposits have been taken from *The Economist, There's oil in them thar sands!* Jun 26th 2003.)

Highly theoretical question to keep you awake at night: Assume that we have 100 years' consumption of oil left – at the present consumption rate. Now, if we increase our consumption by 1% per year yet manage to increase our efficiency in oil usage by 2% each year; how long before we run out of oil? Send me some oil futures for the answer...no, it's in the footnote below.[4]

Applications of price elasticity of supply

Many *primary goods* such as iron ore, teak wood and agricultural goods will be highly inelastic, as the ability to switch to such goods in the short run is very limited. The basic premise here is that once suppliers have a large amount of resources sunk in a certain sector, the re-applicability of these resources is the key to increasing supply of other goods. Just consider commodities such as iron and copper; if you have a copper mine and demand for iron increases it's going to be a bit difficult to switch production to iron.[5]

Secondary goods tend to have higher price elasticity of supply than primary goods. The reasons are simply that secondary goods can be stored, are often produced in industries with a degree of excess capacity, are subject to technological advances in production and are increasingly global in span so that production is close to relevant markets.

---

4   The answer is: Never! (But I'm still bribable. It's the only thing I have in common with Mexican cops.)
5   I mean, you have a *copper* mine…

Summary & revision

1. Price elasticity of supply (PES) shows the responsiveness of producers to a change in price.

2. The value of PES is given by the relative change in quantity supplied over the relative change in price. The formula for PES is

$$PES = \frac{\%\Delta \text{ in Qs}}{\%\Delta \text{ in price}} \rightarrow \frac{\dfrac{Q_1 - Q_0}{Q_0} \times 100}{\dfrac{P_1 - P_0}{P_0} \times 100}$$

3. The range of values for PES is:
   a. PES > 1; supply is elastic
   b. PES < 1; supply is inelastic
   c. PES = 1; supply is unit elastic

4. Any linear supply curve with a P-axis intercept has a PES greater that 1; any linear curve intercepting the origin is unit elastic; any linear supply curve intercepting the Q-axis will have a PES value of less than 1 (inelastic).

PES$_a$: %ΔP < %ΔQs – elastic supply

PES$_b$: %ΔP = %ΔQs – unit elastic supply

PES$_c$: %ΔP > %ΔQs – inelastic supply

5. Determinants of PES are the time span, the storability of the good, excess capacity of producers, closeness of producer substitutes and the ease of attaining more factors of production within the given time period.
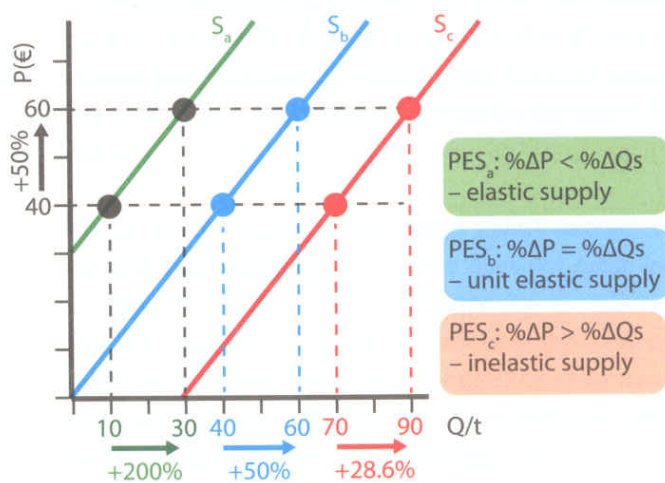
S_0: Immediate supply: non-storable, no excess capacity, no close producer substitutes, no factor mobility. S_1: SR, degree of storability, some excess capacity, some producer substitutes, degree of factor mobility. S_2: LR, storable good, considerable excess capacity, very close producer substitutes, high factor mobility. S_3: VLR, perfectly storable, unlimited capacity, perfect producer substitutes, perfect factor mobility.

6. Primary goods tend to have relatively low PES while secondary (and tertiary) goods tend to be more elastic in supply.

1.3

13. Government Intervention

Key concepts:

- Why governments tax goods
- Specific tax (per unit tax)
- Ad valorem tax

HL extensions:

- Incidence of tax; PED and government revenue
- Incidence of tax; PEDs and government revenue
- Plotting linear supply and demand curves
  - Consumer and supplier surplus
  - Deadweight loss

Section 1.3, Chapters 13 – 15, deals with the various methods governments use to influence supply and demand on markets. We start off by looking at taxes on expenditure.

Why governments tax goods

I always ask my students the question my micro teacher at university asked us: "Why do we tax alcohol more than we tax ice cream?!" There are two answers. The first is that governments levy (= put, apply) indirect taxes as a method to create *tax revenue*. Alcohol will have far lower price elasticity of demand than ice cream so quantity demanded will fall proportionately less for alcohol than ice cream, which means government revenue will be greater. The second reason is that government commonly try to dissuade (= discourage) the consumption of goods which cause damage to society (see Chapter 17). Alcohol causes a great deal of *negative effects for non-users*, such as accidents due to drunk driving and domestic violence.

Specific tax (per unit tax)

A **specific tax** on expenditure is a tax on *each unit* sold/ consumed – which is why such taxes are often referred to as *unit taxes*. This means that the rate of tax per unit is independent of quantity or price. The most common unit taxes are taxes on petrol (€ per litre); on alcohol (€ per litre); and tobacco (€ per kilo or packet). Import taxes are also frequently unit taxes, such as € per tonne of wheat/ per automobile/ per cubic metre of liquid gas. Unit taxes are thus flat rate taxes, which refer to the tax being the same on each unit. A €2 unit tax on a bottle of wine is the same tax on both a €5 bottle of Cote du Merde as on a €50 bottle of 1964 Rioja Gran Reserva.

> **Definition: 'Specific tax'**
> A **specific tax** is the same amount on each unit sold, such as an amount per litre or kilo. It is therefore often called a unit tax.

Assume that government levies an expenditure tax of €2 on, say, cotton shirts. This method of taxation is basically the same as government telling producers that at whatever price they sell the shirts, for each unit sold €2 must go to the tax authorities. Assume that the original retail price (= final consumer price) of an Alligator Shirt is €10 and that the seller buys the shirts for €5. Now, the question is, the suppliers know that however many shirts they sell, each one represents a 'tax debt' to the government – so by how much do they raise the price in order to pay this coming debt?!

Theoretically, producers could still sell the shirts at €10 and simply be satisfied with a gain (*gross profit* in accounting-speak) of €3 per shirt rather than the previous €5. Or, they could raise the price by the same amount as the tax, to €12, in order to keep the same profit margin. The main issue here is how *consumers* react – e.g. their sensitivity to an increase in price. The responsibility of the final price is in the hands of the producer, which is basically the producer thinking "How much of the tax dare I levy on the consumer?!" or "How much should I allow my costs to increase by?" This increase in costs for producers shifts the supply curve to the left.
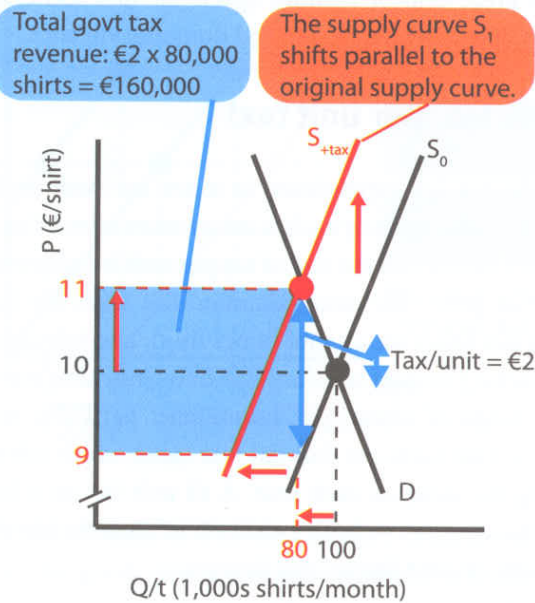
Figure 13.1 a A specific tax on shirts

Figure 13.1 shows how the €2 expenditure tax decreases supply (S_0 to S_{+tax}) by the amount of the tax. The tax can be seen by the vertical distance between S_0 and S_{+tax} – at every level, the difference between supply curves is €2. The tax has the following effects:

- The *price* increases (form €10 to €11) and the *quantity demanded* falls (from 100,000 shirts to 80,000 shirts)

- Half of the €2 tax is paid by the consumer and half is paid by the producer. (This is the incidence of tax – a HL concept done further on.)

- Total government *tax revenue* is €2 × 80,000 shirts; €160,000

- Producer *revenue* decreases from €1 million (€10 × 100,000) to €720,000 (€9 x 80,000) per month

- *Consumption* decreases by 20,000 shirts per month and consumer spending goes from €1 million (€10 × 100,000) to €880,000 (€11 × 80,000).

OUTSIDE THE BOX

My home country of Sweden has the highest overall tax pressure in the world at around 54% of GDP. Income tax is a major part of government revenue and of course a burden on households. My Swedish students were well aware of the incentive for parents to hire black market labour in, say, re-doing the kitchen or adding on to the garage. Several asked during the first edition of this book that I include in the next edition what I did on the board in class, namely illustrate the incentive for me to join the black market for labour during a summer when I was a poor student.* (Note that while I am using the diagram below to illustrate a personal story, it is equally re-applicable to the aggregate, i.e. the labour market as a whole.)

Figure 13.1 b

Basically what happened was that a friend of mine who ran a large retail store needed help one summer stocking goods on shelves and packaging goods for customers. (SEK = Swedish crowns – and I don't remember the wages so I've indexed the va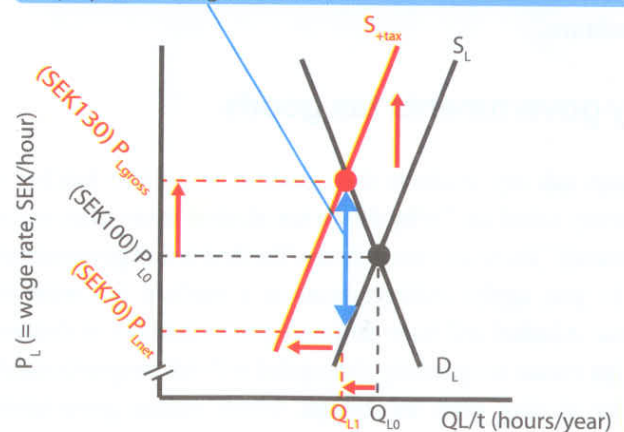lues.) "Look Matt, if I pay you officially, then I'll be subject to labour taxes and social security payments. Your wage of SEK100 per hour will cost me an additional SEK30. On the SEK100 you will pay 30% in income tax. In other words; your net wage of SEK70 will cost me SEK130. Now, what if we split the difference and cut out the 'middleman' in the government tax office…"

It didn't take a degree in rocket surgery to figure out that by avoiding the taxman, my employer could save SEK30 and I could increase my wage by SEK30. I would be willing to work more hours (Q_{L0}) and he would be willing to pay for it. Everybody wins…except, of course, the tax office. Oh, and the people who benefit from public and merit goods; 9 million Swedes.

* To anybody in the Swedish tax dictatorship bureaucracy who happens to see this; we are well past the statute of limitations! Just give me my pension fund and I'll be on my way.

Ad valorem tax

A tax based on the base value of a good is just like any percentage-based tax. For example, a 30% income tax means you pay £30 on an income of £100, £60 on an income of £200 and so forth. The same effect is evident when levying an *ad valorem tax* (value-added tax or VAT); the tax per unit will increase as the price of the good rises.

> **Definition: 'Ad valorem tax'**
>
> An *ad valorem tax* (value added tax) is based on the base value (price) of a good sold, and as it is a percentage the amount of tax per unit will increase as the base value increases.

Figure 13.2 shows the effect of VAT on our Alligator Shirts. Assume that a VAT of 20% is imposed at an equilibrium price of €11 and a quantity of 80,000. The supply curve shifts *disproportionally*, from S_0 to S_{+VAT}, since the higher the price, the larger the price increase of the 20% VAT.
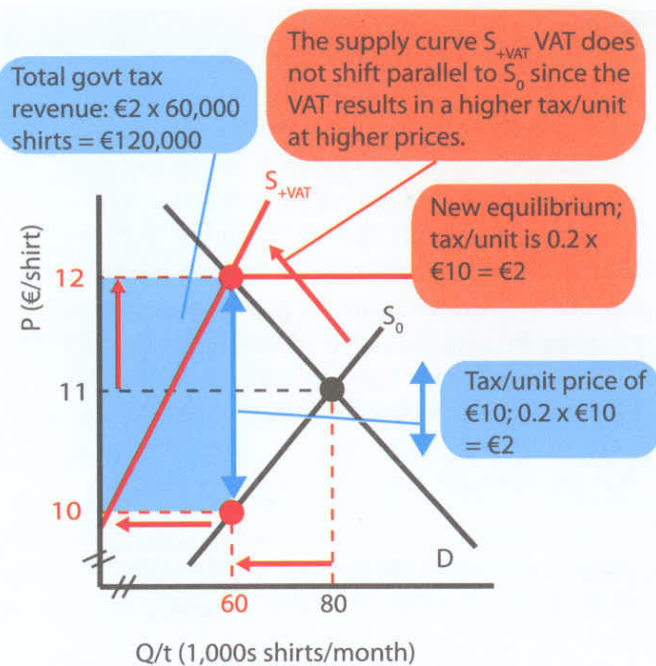
Figure 13.2 Incidence of an ad valorem (VAT) expenditure tax on shirts

At higher prices the ad valorem tax shifts the supply curve higher, increasing the distance between S_0 and S_{+vat}. At the new equilibrium price of €12, the 20% VAT renders a €2 tax per unit.[1] In other respects we have the same type of outcome as before; the price increases from €11 to €12, quantity demanded decreases from 80,000 to 60,000 shirts, producers' revenue is €600,000 (€10 × 60,000) and consumer spending is €720,000 (€12 × 60,000).

Exam tip; clarity in diagrams

Note in Figures 13.1 and 13.2 how I 'break out' the tax arrow and refer to it clearly as 'Tax/unit'. When we add in additional economic concepts, such as welfare loss triangles (Chapters 17 and 18), you will make your examiner's life easier by keeping all points, arrows and distances clearly distinguishable. Basically, as I tell all my people; "Do you want to yank the chain of the person grading your exams?"

1 A brief footnote on my examples above (Figures 13.1 and 13.2). Many of you are probably wondering what lunacy possessed me to put first a unit tax – and then a value added tax on top of that! The example of 'taxing a tax' looks absolutely bizarre, silly and impractical. Quite naturally, this is exactly what is done in many countries. Petrol prices in Europe, for example, will be comprised of around 80% tax. Say that the base price of a litre of petrol is €0.2 to which a €0.6 unit tax (per litre) is added, and then topped off by a 25% VAT. This brings the final price to €0.8 + 25% = €1.

I recommend that you do the 'break-out the tax arrow' (or something like it) in all your diagrams and basically make it a 'knee-jerk' (= natural) reaction. During exams you will be stressed out and that's when you get messy/unclear – and lose marks as a result. One of my karate teachers, Kawasoe sensei, was very clear that what you practice in the dojo (= karate gym) is what you do in a stressful situation (= getting attacked*). He was right. I am asking you to make neat and clear diagrams a part of your "econ karate nature" so that when the exam questions assault you, your neatness-reaction is second nature.

* You think I'm making this up? Listen, I live on coffee, sugar, nicotine and stress. Basically I'm pretty wired. When I'm was out running in "bad areas" in Mexico I was frequently attacked by wild dogs – carrying any number of diseases which are probably not even in the books. To date I've palm-smacked and front-kicked at least three dogs into the brush. I sent a mental note of thanks to Kawasoe sensei every time. Oh, after a while, fearing getting some nasty disease via physical contact with the mutts, I had the workers in school make me a nice pair of nunchuks. I got more exercise swinging the nunchuks at dogs than from the actual running.

A Case (of stupidity?) Study;

Tobacco tax in Sweden1997/98

Some movies should be obligatory viewing for politicians. The Swedish government would have done well to watch "The Untouchables" (about the infamous Al Capone and organised crime's rise to power during the ban on alcohol in America in the 1930s) prior to increasing tobacco taxes in 1997. Here's what happened.

In August of 1997, the Swedish government raised the tax on tobacco by 18% as part of a health-drive and possibly to increase tax revenue. Nobody actually seems to know. In any case, cigarette consumption plummeted by 20% which would seem to belie the 0.45 PED for tobacco in Table 9.1 in Chapter 9.

What the Swedish government did not seem to realise, yet every single one of my IB1 students did, was that there were indeed any number of substitutes available for cigarettes which would explain the massive 20% fall in consumption, pointing instead to a PED of 1.1. It turned out that the Baltic states, Russia and Poland had an abundance of cheap black market cigarettes and also a number of budding capitalists who saw an opportunity to make a good deal of money. "Duh!" said all my IB1 students who snidely predicted exactly what was to happen!

Figure 13.3: The increase in tax lowered the legal quantity demanded from Q_{97} to Q_{98} raising the official market (O.M.) price of cigarettes from P_{97} (point A) to P_{98} (point B).

At each price level, the black market price is lower than the official market price.
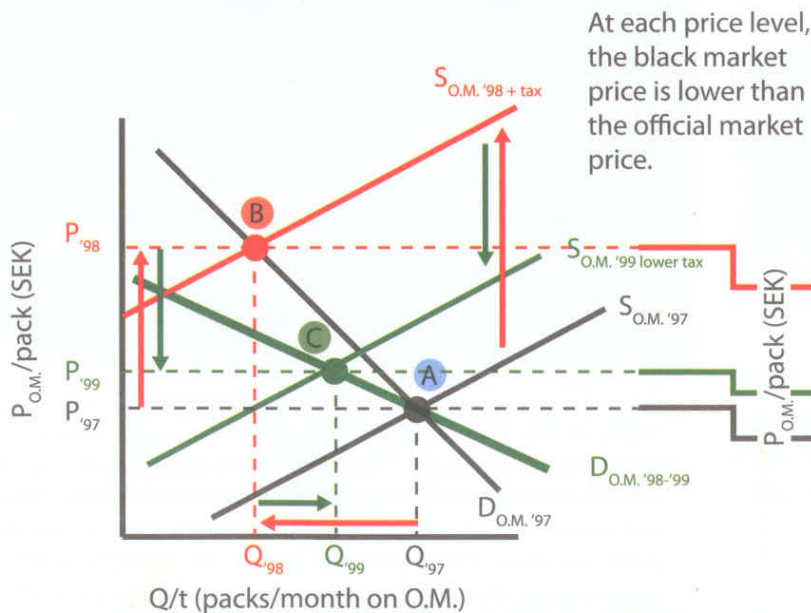
Figure 13.3 Cigarettes – official market (O.M.)
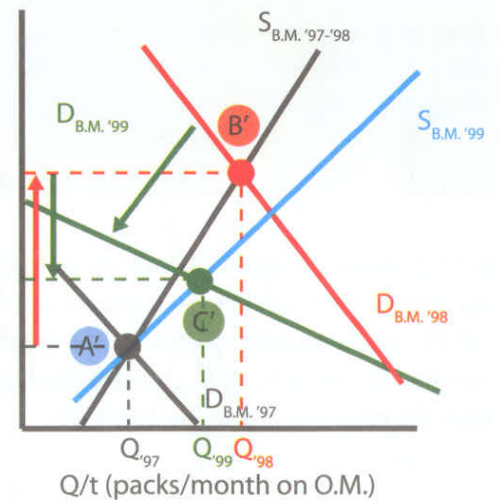
Figure 13.4 Cigarettes – black market (B.M.)

Figure 13.4: An immediate surge in demand for black market cigarettes was the result, where demand rose from $D_{B.M.97}$ to $D_{B.M.98}$ and quantity went from Q_{97} to Q_{98} (point B').

Secondary effects: During the period of 1997/98, demand for official market cigarettes had become more elastic, due to the availability of substitutes. By the time the tax on cigarettes was lowered, demand had decreased to $D_{OM\ '98-'99}$ and become more price elastic due to the increase in available substitutes. The effect was that quantity demanded didn't return to the previous level, but only to Q99 (point C). The reason for this was that the supply of black market cigarettes had increased markedly, from $S_{B.M.'97-'98}$ to $S_{B.M.99}$. Black market PES increased also, since by this time the black marketers' supply chains and stocks had been firmly established. The lower tax did have an effect on demand for illegal cigarettes, as shown by the shift of $D_{B.M.\ '98}$ to $D_{B.M.\ '99}$, yet the black market equilibrium did not return to the previous level (point B') but was established at point C'. Throughout the period, the black market price was kept firmly below the official price level, making it possible for black marketers to remain entrenched (= established) on the market.
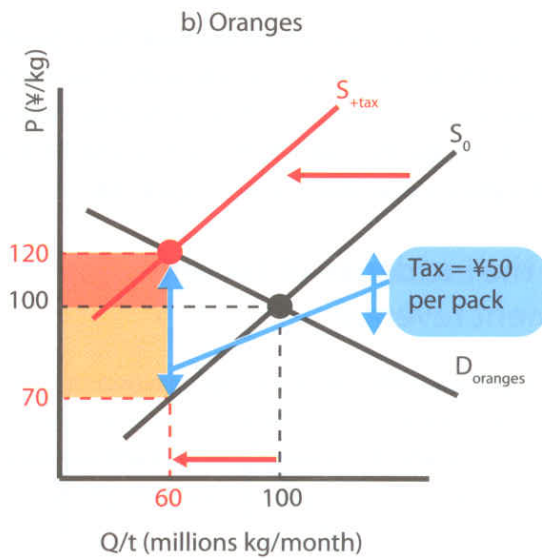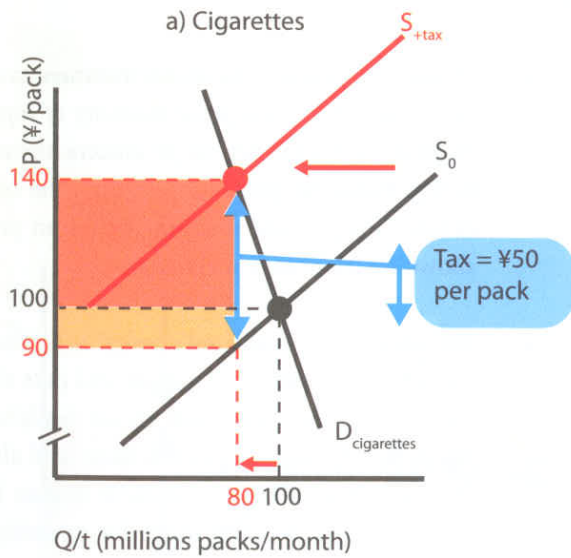
HL Extensions

PED – the incidence of tax and government revenue

Let us compare two goods, one with low PED and one with high PED; cigarettes and oranges. This time, we extend the analysis from how consumers and producers will be affected by the tax to how government revenue will differ. Figure 13.5 below shows supply and demand for the two goods, and we assume rather unrealistically both goods to have an initial price of ¥100 (Japanese Yen) and quantity of 100 (basically, I am indexing prices and quantity). A specific tax of ¥50 is put on both goods,[2] raising the price of both goods – but the comparison pretty much ends there.

_____

2   I will limit myself to flat rate taxes in showing incidences of tax. There is simply no real benefit in using an ad valorem tax to show the burden of tax.

Figure 13.5   Expenditure tax on a) Cigarettes & b) Oranges

111

Both goods' demand curves will have a marked effect upon whom the main incidence of tax will fall. The lower the PED, the higher the increase in price due to the expenditure tax – and the more of the incidence of tax lands upon the consumer.

- Incidence of tax on the consumer = $\dfrac{\Delta \text{ in } P}{T}$ where P is the price to the consumer and T is the tax per unit on the good. The proportional incidence of cigarette tax on the consumer is thus 40/50 which is 80%.

- Incidence of tax on the producer = $1 - \dfrac{\Delta \text{ in } P}{T}$ which is 1 – 0.8 or 20%

The unit tax on cigarettes (Figure 13.5a) of ¥50 caused a far higher price hike than for oranges and put 80% of the *incidence of tax on the consumer*. This makes intuitive sense, for as soon as consumers lack viable alternatives they will be willing to pay a great deal of a price increase, something that firms are well aware of. Of the total incidence of tax, consumers bear the burden of ¥40 × 80 million = ¥3,200 million while producers bear the remaining ¥800 million.

Oranges (Figure 13.5b) on the other hand have many possible substitutes – and hence a higher price elasticity of demand – and this shows since *consumers only bear 20% of the total tax burden*. As for the question of which good would be more suitable for taxation, I leave that to you to figure out.

PES and the incidence of tax and government revenue

We now look at the incidences of indirect taxation in relation to different *supply* curves. Let me be a trifle provocative in the choice of goods with which to exemplify this. I will use nuclear energy as one good and coal as the other. Let's put in a few assumptions:

1. Nuclear power and coal are the only two available sources of electricity;

2. Both are perfect consumer substitutes with identical PED values;

3. Both industries are operating close to maximum output; and

4. Increasing nuclear power is far more supply inelastic than increasing coal powered electricity, as it takes a great deal more time to install additional nuclear reactors than coal furnaces.

Once again, we assume that the original price and quantity are the same for both goods. Government now looks at the possibility of levying a 'green tax' on one of the two alternatives. (I forego using actual figures now as you are assumed to be on-line.)

- Figure 13.6a shows how an environment tax on nuclear power which has *low price elasticity of supply* will put most of the tax burden on producers; the shift from S_0 to S_{+tax} raises the price far less than the total amount of tax. This puts the main incidence on producers, as shown by the orange rectangle.

- The alternative, coal-fired energy in Figure 13.6b, has higher price elasticity of supply and puts the *main tax burden on consumers*. The choice boils down, once again, to 'who should pay the most' and also which of the two alternatives government wishes to dissuade use of. It is also clearly evident that quantity demanded decreases more due to a tax when price elasticity of supply is high.

The final four examples show how PES and PED affect the incidence of tax such that it is 100% on either the producer or consumer.

a) Nuclear power   b) Coal power

Incidence of tax on producer   Incidence of tax on consumer
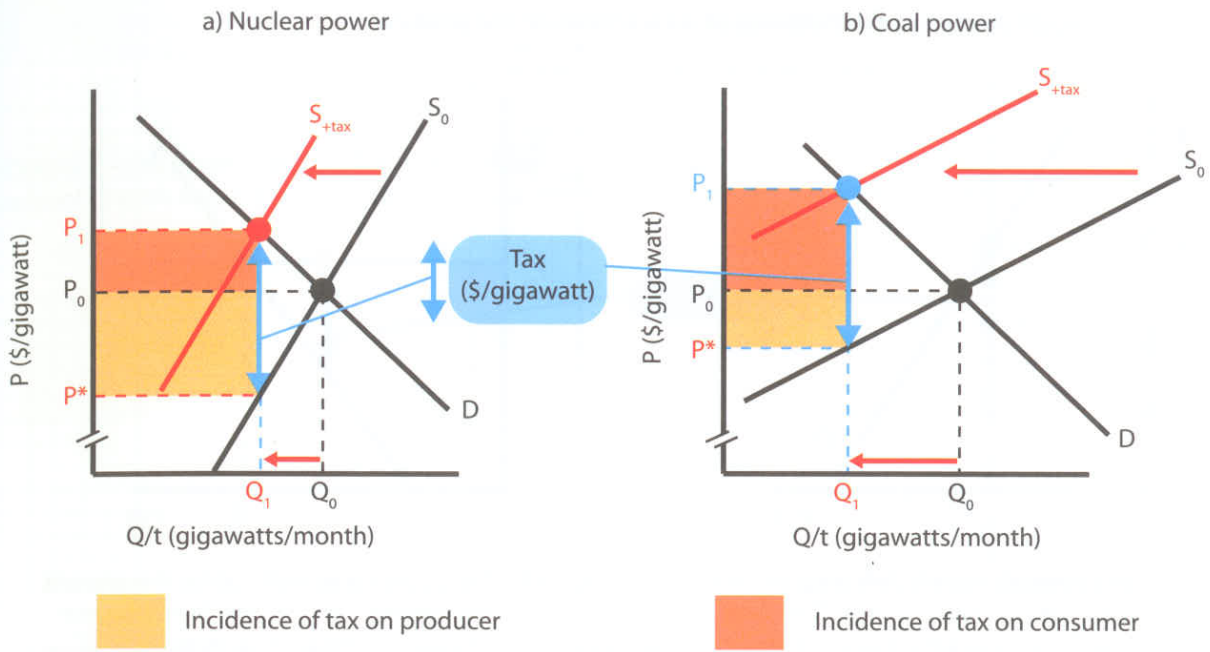
Figure 13.6a) & b)   Effect of an expenditure tax on nuclear and coal power

$\Delta P = tax/unit$; the incidence of tax is entirely on the *consumer*
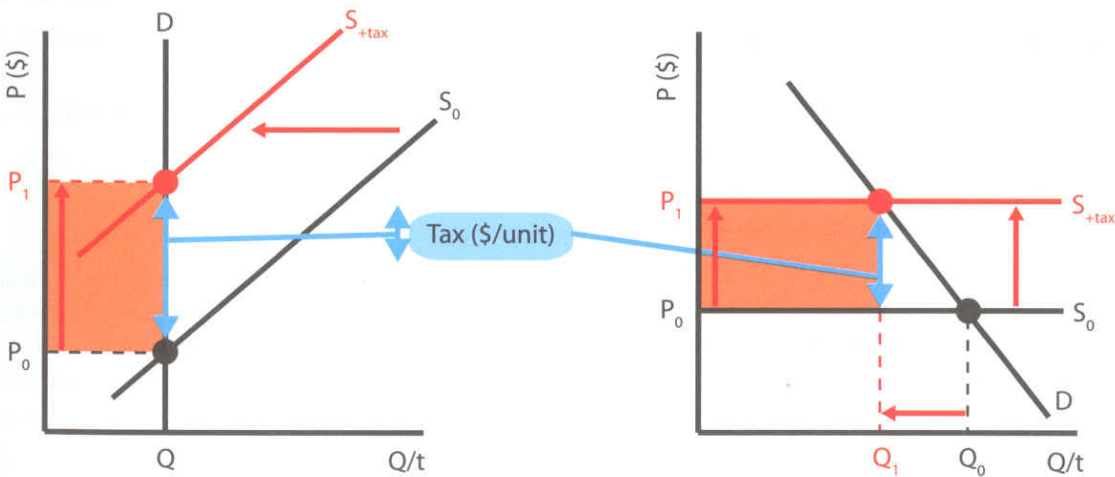
Perfectly inelastic demand (left diagram) will raise the price by the same amount of the tax (no change in Q_d), as will perfectly elastic supply (right diagram). The entire incidence of tax thereby falls on the consumer.

No ΔP due to tax; the incidence of tax is entirely on the *producer*
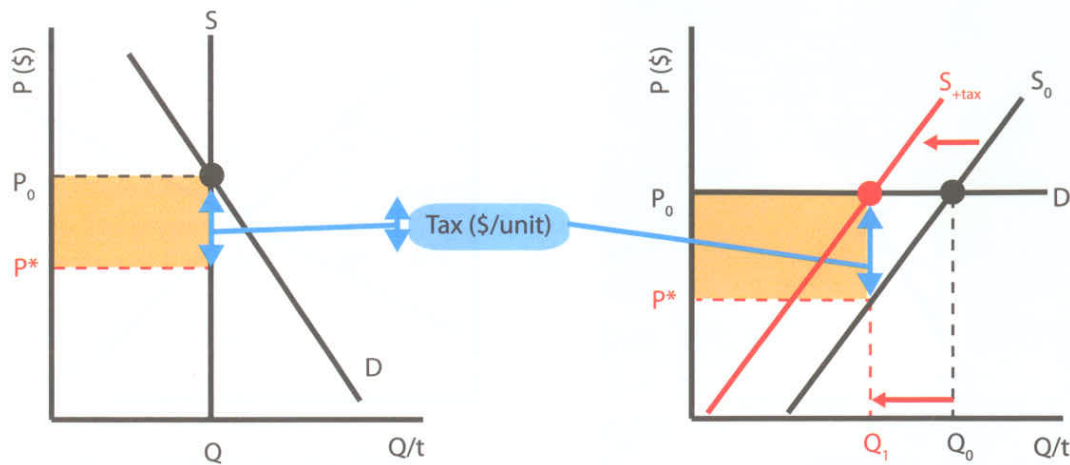
Perfectly inelastic supply (left diagram) is trickier. Technically, the supply curve shifts up by the amount of the unit tax, i.e. stays in place. There is no change in quantity demanded – any tax will have to come out of producers' pockets. A perfectly elastic demand curve (right diagram) means that demand is constant at a given price. Any movement along the demand means the price remains the same, so the entire incidence of tax is levied on the producer.

Figure 13.7 (a) & (b) Four extreme cases of the incidence of tax

Hint: Don't try to commit the above to memory. I still haven't! All one needs to do is remember that if a tax is levied and the price changes by the same amount as the tax, the consumer will bear the entire tax burden. In a case where the price doesn't change, then the producer will bear the entire incidence of tax.

Plotting linear supply and demand curves

Here we will use the previous supply and demand functions from Chapters 4 – 6 to show and calculate the effects of an expenditure tax on goods. You will have to forgive me for including the 'invisible elongated' portion of the Q-axis (i.e. the section showing minus values) for reasons of explanatory clarity. For the same reason, I have put a rather hefty tax on this particular good – since a $2 tax would be rather hard to see clearly in a diagram where the starting value is $100.

So, with a equilibrium price and quantity of $100 and 2,000 units, we add a $50 tax:

- The **supply function** in $Qs = -2,000 + 40P$

- The new supply function needs to be calculated. A $50 tax raises the supply curve S_0 by $50 at all quantity levels, thus we need to calculate **the new value of 'c'**.

  o The **original P-intercept** of the supply curve is 50

  o The P-intercept of the supply curve is calculated by 'c' / 'd'

  o We know that the **new P-intercept is 100** and that the slope ('d') is unchanged – thus we can calculate 'c' by solving $100 = -c / 40$; $-c = 40 \times 100$

  o The new value of 'c' is -4,000

- The new supply function is $Qs = -4,000 + 40P$

Now is a good time to refresh your memory by calculating the new equilibrium price and quantity. These values have been left blank in Figure 13.8 for precisely that reason.

The slope of P with respect to Q is $50 \div 1,000 = 1/40$
The slope of Q with respect to P is $1,000 \div 50 = 40$

$Qs = -4000 + 40P$ or $P = 100 + 1/40 Qs$

$Qs = -2000 + 40P$ or $P = 50 + 1/40 Qs$

$Qd = 4000 - 20P$ or $P = 200 - 1/20 Qd$

$P_2 = ?$

$P_1 = ?$

+50

+1,000

$Q = ?$
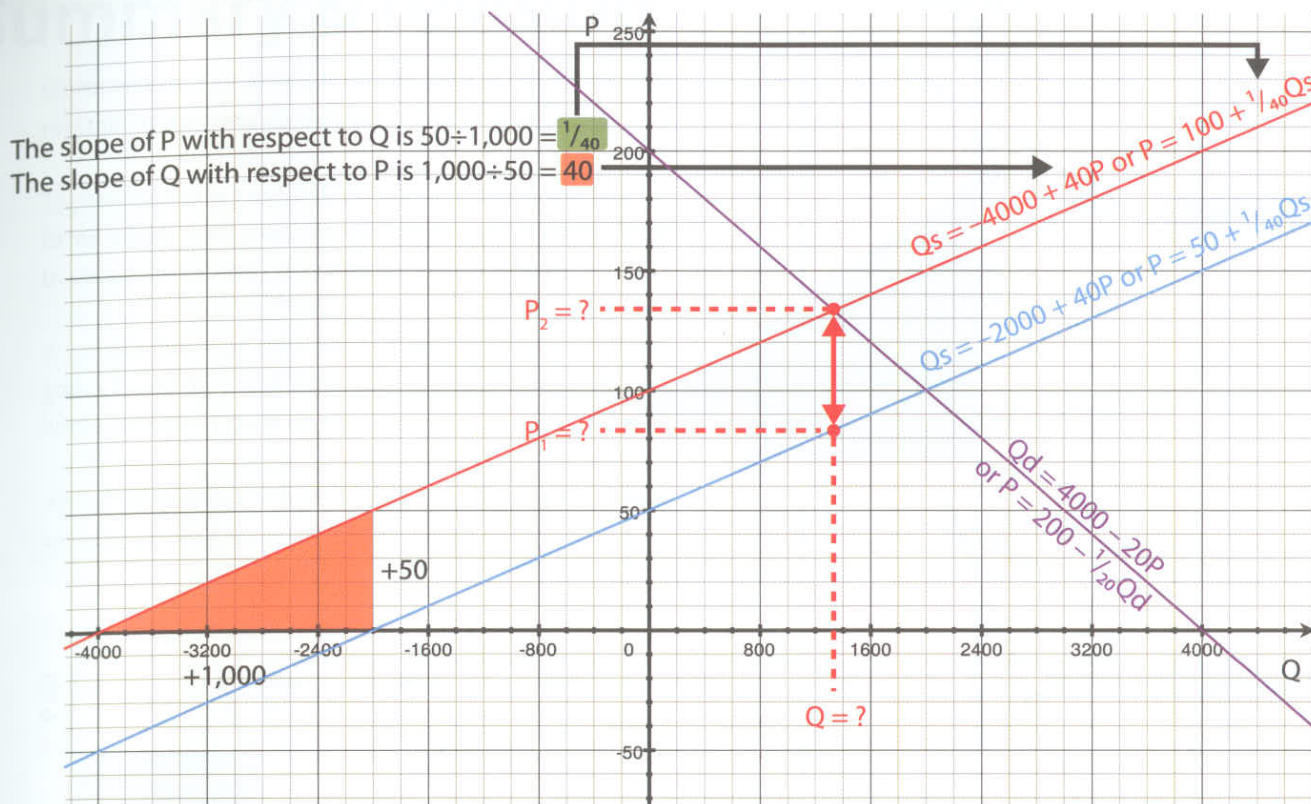
Figure 13.8 Unit tax on a good

New equilibrium quantity is given by solving the simultaneous equation $-4000 + 40P = 4000 - 20P$

- P; $8,000 = 60P$, P = **$133.33**

- Qs; $-4,000 + 40 \times 133.33 =$ **1,333**

Incidence of tax

There is an initial clue here that will help you check your figures! Any case where the slope of a (linear) supply curve is less than the slope of a (linear) demand curve, we know who will carry the largest incidence of tax; the consumer. Doodle a few diagrams to see why. (All 'areas' in the text below refer to Figure 13.9).

- Total incidence (areas 2 and 4); at a quantity of 1,333 and tax of $50, the total incidence (e.g. total *government revenue* is **$66,666**.

- Incidence of tax on *consumers*; the increase in price is $33.33. Consumers will pay $33.33 × 1,333 in tax which is **$44,428**.

- *Suppliers* will pay the remaining **$22,217**.

Consumer and supplier surplus

Go back and revise consumer surplus, supplier surplus and deadweight loss. That will make it easier for you to see the areas in Figure 13.9 and follow the calculations. My wonderful classroom neighbour, Brett the Aussie Math Teacher and Ballroom Dancer, tells me that the easiest way to calculate remaining surplus is "…quantity times the relevant P-intercept minus everything else … then divide by two …" Or something like that.

- Remaining *consumer surplus* (area 1):

$$\frac{(\$200 - \$133.33) \times 1333}{2} = \$44,435.5$$

- Remaining *supplier surplus* (area 6):

$$\frac{(83.33 - 50) \times 1333}{2} = \$22,214.5$$
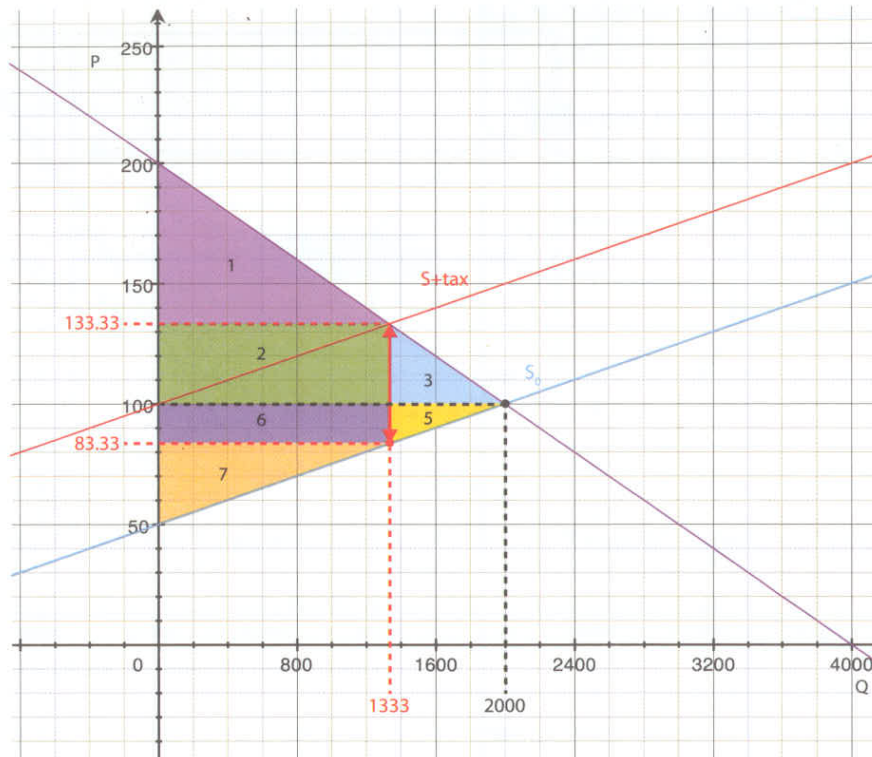
Figure 13.9 Consumer and supplier surplus, incidence of tax and deadweight loss

Deadweight loss

The deadweight loss is the sum of the net loss of consumer and supplier surplus. These are the two small triangles (areas 3 and 5 in Figure 13.9) that are simply 'lost' in levying a tax on a good. Looking at the $50 unit tax as the 'base of a triangle', we get:

$$\frac{\$50 \times (2,000 - 1,333)}{2} = \$16,675.$$

POPQUIZ 13.1

1. Assume a supply function of $Qs = 200 + 10P$ and a demand function of $Qd = 500 - 5P$.

2. Draw a diagram showing initial equilibrium.

3. Add a unit expenditure tax of $10 and calculate the new supply function. Draw a new supply curve showing equilibrium price and quantity after the tax.

4. Calculate the total incidence of tax (government tax revenue).

5. Calculate the loss of consumer and supplier surplus.

6. Calculate the deadweight loss.

Summary & revision

1. Governments tax goods to either gain government tax revenue or decrease consumption of the good.

2. Expenditure taxes are *indirect taxes*. Consumers pay to firms which then pass on the tax to government. This increases firms' costs and thus decreases supply.

3. A *specific tax* – or unit tax – is an expenditure tax on goods based on a unit sold or consumed such as €2 per bottle of wine or pack of cigarettes.

4. An *ad valorem tax* is a percentage added on to the base value or price of the good.

High PED

Low PED

Specific tax (unit tax)

Ad valorem tax

Incidence of tax on producer
Incidence of tax on consumer

HL extensions

5. Assume a supply function of **Qs = -400 + 20 P**. The P-intercept for the function is given by 'c' / 'd'; 400/20 = $20. A unit tax of $5 will shift the supply curve upwards by the same amount – the P-intercept for the supply curve would go from $20 to $25.
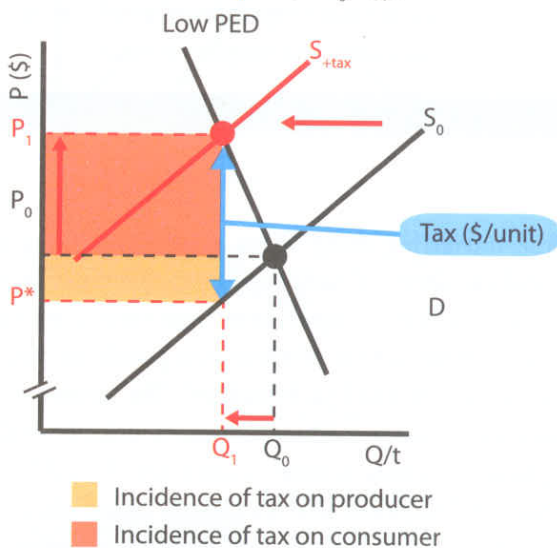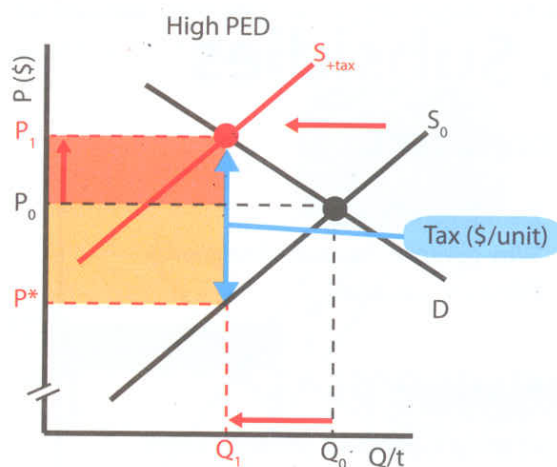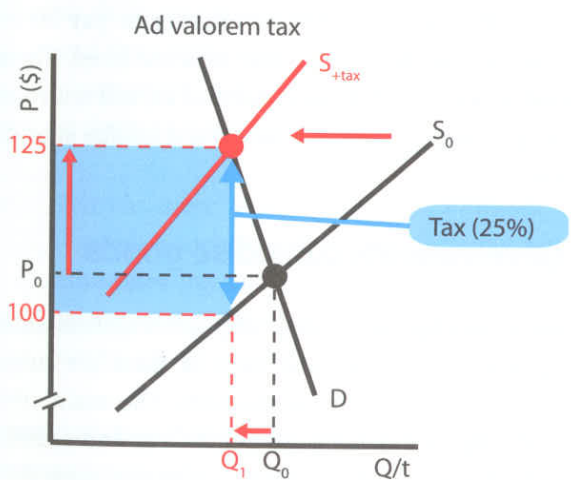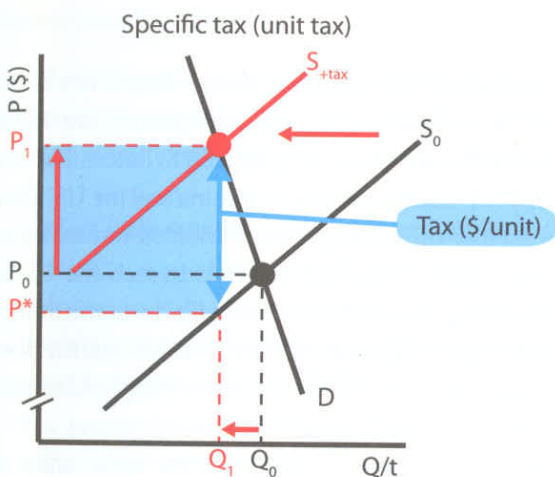
6. To get the **new supply function**, we need to calculate the new value of 'c'. This is given by solving 'c' in the equation -c/20 = 25. The new value of 'c' is -500. Another way to calculate 'c' is by shifting the supply curve to the left by the same proportion as the tax; the $5 tax is an increase of 25% so the supply curve will intercept the Q-axis at -400 × 1.25, i.e. -500.

7. **Total incidence** of tax is calculated by multiplying the unit tax by the new equilibrium quantity.

   a. Incidence of tax on the **consumer**; (new equilibrium price minus original price) times new equilibrium quantity

   b. Incidence of tax on the **producer**; total incidence minus the incidence on consumers

8. **Deadweight loss;** $\dfrac{(\text{tax per unit} \times (Q_0 - Q_1)}{2}$

14. Subsidies

Key concepts:

- Definition of subsidies
- Why governments subsidise goods
- Effects of subsidies

HL extensions:

- Plotting supply and demand curves
  - Government expenditure, producer revenue and consumer spending
  - Consumer and supplier surplus

The interesting thing about economics is the almost endless array of seeming contradictions. Recall the 'guns or butter' opportunity cost issue I brought up in Section 1. Now, which of these two goods do you think governments are inclined to subsidise?! The answer is 'Both'. Governments consider military defence a priority and often render grants to private military contractors for R&D. As for butter, rural development and traditional values often lead to farm subsidies so that farmers can attain what is considered reasonable income levels.

Definition of subsidies

In any case, a direct **subsidy** is a money payment to producers in order to increase the supply of the good and lower the price. This gift or grant from government increases supply for two reasons:

1. The subsidy acts as an incentive for producers to produce more

2. The subsidy serves to lower (marginal) costs of production

---

Definition: 'Subsidy'

A **subsidy** is a money gift/grant from government which acts as a) an incentive to producers to produce more; and b) lowers the (marginal) costs of production. Both of these forces serve to increase supply.

Notice the use of 'direct subsidy' above! This is not accidental; many of the subsidies given by government are hidden or implicit subsidies which are very difficult to detect. For example, in the 2010-'11 'trade war' between China and the US, China has been accused of subsidising export oriented industries via low interest loans and cheap land to firms. In fact, the US is often accused of doing the exact same thing for domestic industries – the World Trade Organisation (WTO) ruled against Boeing in 2010 for violations of WTO rules concerning hidden subsidies such as tax breaks and research and development aid.[1] In my home country of Sweden, governments have often sought to lower youth unemployment by offering to pay 30 – 50% of firms' wage costs if 18 to 22 year olds are hired. There is a lengthy list of such 'hidden' subsidies and we will return to this highly inflamed international debate about hidden subsidies in Chapter 65.

Why governments subsidise goods

There are a number of reasons why governments subsidise certain goods or industries and most of them are subject to some serious criticism from economists. The main criticism is that subsidies are a form of hidden tax-and-redistribute policies which lower allocative efficiency by distorting the price mechanism.

The main reasons for subsidies are:

- Increasing *employment* – increased output in firms can create demand for labour and unemployment falls (infrastructure projects such as bridges and dams)

---

1   Washington Post, *US subsidies violate trade law*, September 15 2010

- Protecting a way of life and *culture* – traditional way of life in farming/rural areas can be upheld (farming subsidies in the EU)

Vietnamese Rice Paddy

- *Redistribution* of income – disadvantaged industries receiving subsidies can retain wage levels for employees (mining industries in the UK in the 1980s)

- Meet *international competition* and improve the trade balance – as subsidies lower the costs of production, domestic industries become more price competitive which can increase exports and/or decrease imports (oil industry subsidies in the US)

- Improve *production* potential and *products* – subsidies to firms and universities serve to increase research and development (R&D) spending which can increase productivity and spur innovation (alternative energy in Spain and Denmark)

- Provide more goods which are *societally beneficial* – goods which have positive effects not only on the consumer but other groups (third parties) are said to be under-consumed without a subsidy (milk and dental care in Sweden)

It bears repeating that there is very little in the way of general agreement amongst economists as to which subsidies have net positive effects on society – e.g. whether the costs outweigh the benefits. This is brought up in Section 1.4.

Effects of subsidies

The European Union's Common Agricultural Policy (CAP) spends some USD50 billion a year on agricultural subsidies. The

effect has thus been periodic over-supply. Figure 14.1 shows an example of how a subsidy on wheat would affect wheat output.

The supply curve $S_{+subsidy}$ shifts parallel to the original curve.

Subsidy/unit = €10
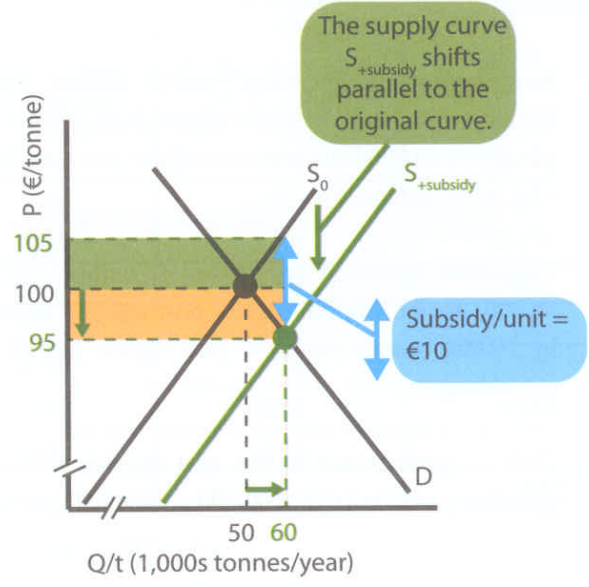
Incidence (= benefit) of subsidy on the *supplier*

Incidence (= benefit) of subsidy on the *consumer*

Total incidence of the subsidy, e.g. govt. cost of the subsidy: €10 x 60,000 tonnes = €600,000

Figure 14.1 Incidence of a subsidy on wheat

Market equilibrium is at a price of €100 per metric tonne and a quantity of €50 million tonnes. When a subsidy of €10 per tonne of wheat is granted, supply increases, S_0 to $S_{+subsidy}$, and a new equilibrium price of €95 is reached. The subsidy increases supply which in turn lowers the price and increases the quantity demanded. The amount supplied on the market has increased to 60 million tonnes. The total incidence of the subsidy is €10 per tonne times 60 million tonnes; €600 million. (Please compare the total incidence of the subsidy in Figure 14.1 with the total incidence of a tax done in Figure 14.2 [effect of a flat rate tax] in Chapter 13. Don't mix the two up – it's a common error.[2])

_____

2   In dealing with total incidences of either tax or subsidy, I always tell my people the following trick: 'Find the new equilibrium point (where $S_1 = D$), draw a line straight up/down till you hit the old supply curve. Then go 'home' – to the P-axis – from both points that you have on the two supply curves. This **square horse-shoe shape** is total incidence of tax or subsidy.'

There are additional effects of a subsidy, namely the increase in producers' revenue and the change in consumers' consumption and spending:

- **Producer revenue** increases from €5 million (€100 × 50,000) to €6,3 million (€10 × 60,000 from government plus €95 × 60,000 from consumers) per year (see Figure 14.2 in HL extension)

- **Consumer spending** goes from €5 million (€100 × 50,000) to €5.7 million and consumption increases by 10,000 tonnes per year. (see Figure 14.3 in HL extension)

- **Cost to government** is the unit subsidy times the quantity produced; €10 × 60,000 = €600,000 per year.

Evaluation of subsidies (over dinner)

There are numerous issues arising in the area of subsidies that can and should be discussed. The main three areas deal with costs, equity and developing countries.

- **Actual costs of subsidies:** I once asked a lady at the milk section in a grocery store in Sweden if she thought milk was expensive – and to my amazement she said "Not as expensive as it would be without the subsidy!" She had her mind screwed on right – just look at Figure 14.1 and calculate the *actual* cost per tonne of wheat! Yes, the entire subsidy from government comes from tax money. Milk buyers are paying for this subsidy indirectly:

  o *Inefficiency* costs arise since the decision to take my tax money and give it to wheat/milk producers is made by the government. It would be far more efficient to let the price mechanism allocate resources.

  o All of the subsidies will entail *administrative* costs.

  o The heaviest costs are of course the *opportunity* costs of the subsidies. Most of us in the developed world could think of a number of better things to do with the money than to give it to farmers.

- **Equity**[3]: I also asked a French ex-wife's family at their dinner table one time "…what do you think about me, the Swede, paying tax money so you French can enjoy cheap cheese?"[4] Looooong silence and then… no answer. The fact is that Sweden is one of the largest *net* contributors to the EU and thus the Common Agricultural Policy (CAP) and France is one of the largest *net* beneficiaries. Even if I had been in Sweden I would have asked the question of a Swedish farmer "…why are farmers considered so valuable…why not cobblers or thatchers …" The basic question here is on what grounds government – not the consumer! – decides who gets subsidies.

- **Developing countries**: I hate to run this into the ground but here's another one. At a dinner party here in Mexico I asked my hosts (who are *still* my friends!) "…so, how do you feel about US subsidies on cotton, corn and pork?" My ears are still burning from the reply, which was along the lines of how US producers were dumping[5] agricultural products on Mexico and destroying domestic markets due to the fall in prices. (Like a good economist, I looked it up. See the table below.) There are numerous examples of how subsidies in high income countries create surpluses of agricultural produce which, when dumped on LDCs cause not only loss of LDC export income but also the destruction of domestic LDC markets. In Chapter 15 I shall return to the issue of how rich world intervention on agricultural markets destroys livelihoods in poor countries.

---

3 Equity means fairness in distribution of societal resources.
4 Yes, I actually did this. Nothing like a little light chit-chat to make a dinner pleasant. Dessert was…chilled, shall we say.
5 This has a very specific definition in economics. It is the selling of goods on foreign markets at below the cost of production at home.

U.S. Dumping on Mexican Producers

	United States		Mexico		
	Exports 90-2 - 06-08 growth %	Dumping margin avg 97-05	Producer Prices 90-2 -05 real pesos	Production volume 90-2 - 06-08	Dumping Losses 1997-2005 SUS millions
Corn	413%	19%	−66%	50%	6,571
Soybeans	159%	12%	−67%	−83%	31
Wheat	599%	34%	−58%	−7%	2,176
Cotton	531%	38%	−65%	−3%	805
Rice	524%	16%	−51%	−8%	67
Beef	278%	5%	−45%	31%	1,566
Pork	707%	10%	−56%	40%	1,161
Poultry	363%	10%	−44%	133%	455
Total					12,832

Sources: USDA-FATUS; IATP; starmer et al. (2006); SAGARPA. Constant 2000 USS.

HL extensions

Plotting supply and demand curves for subsidy

The methodology here is very similar to that of calculating the incidence of tax. A subsidy will shift the supply curve parallel to the old and you will have to be able to calculate the new supply function, new equilibrium and the resulting changes in producers' revenue, consumers' expenditure, government costs, societal surplus and deadweight loss.

Same old thing again; we assume a supply and demand function on the market for corn. $Qd = 1,100 - 25P$ and $Qs = -900 + 25P$. (By now you shrewdly notice that the slopes of the demand and supply curves are the same. What does this mean for the incidence of the subsidy?[6]) Government now levies a subsidy of €6 per tonne as illustrated in Figure 14.2.

Government expenditure, producer revenue and consumer spending

The €6 subsidy will shift the supply curve to the right. We start, as usual, by calculating the new supply function and then the new equilibrium. Note that the price and quantity axes in Figure 14.2 have been broken in order to clarify the relevant areas.

- The original supply function is $Qs = -900 + 25P$ and the P-intercept ('c' / 'd') of the supply curve is €36.

  - The new P-intercept for the supply curve will be €30

  - Solving $30 = \frac{-c}{25}$ gives c = -750

  - The new supply function is $Qs = -750 + 25P$

a) Producer revenue

Subsidy = €6 per tonne

S_0

$S_{+subsidy}$

43

40

37

D

100   175

Q/t (1,000s tonnes /month)

P (€/tonne)

43

40

37

Cost to govt (= total incidence of subsidy)
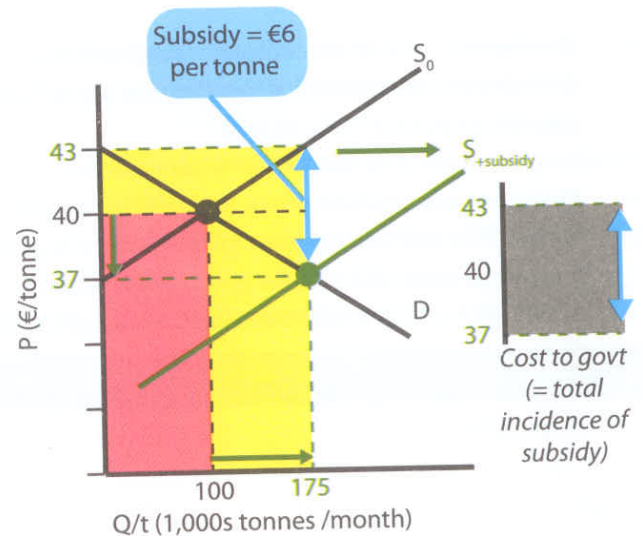
Producers' revenue before subsidy

Producers' *increase in revenue* due to subsidy

---

6   It means that the incidence of the subsidy will be divided equally between consumers and suppliers.

b) Consumer spending

Subsidy = €6 per tonne

Consumers' *decrease in spending* due to subsidy

Consumers' *increase in spending* due to subsidy
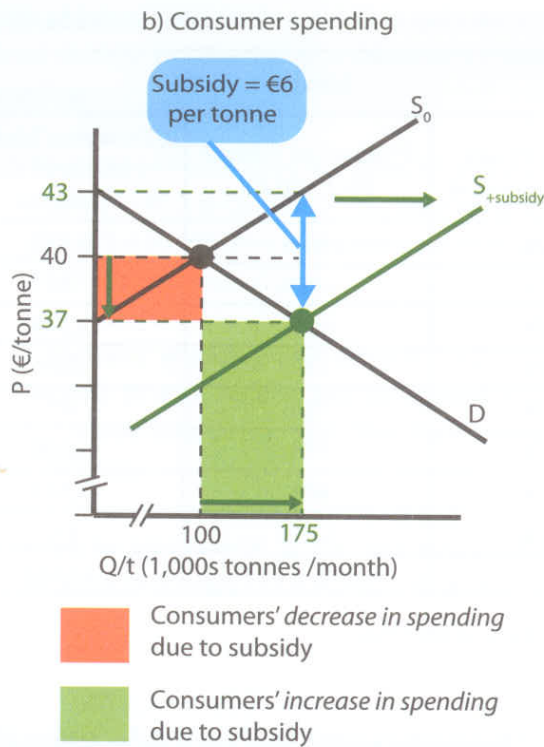
Figure 14.2 Effects of a subsidy on corn

New equilibrium price and quantity are given by solving the simultaneous equation **1,100 − 25P = -750 + 25P**

- P; 1,850 = 50P, P = **€37**

- Qs; -750 + 25 × 37 = **175 tonnes**

The cost and re-distribution issues are as follows:

- *Government cost* of the subsidy (total incidence of the subsidy); the subsidy times the new equilibrium quantity gives €6 × 175 = €1,050

- *Producers' gain* in total revenue; initial revenue is €4,000 (€40 x 100) and total revenue post-subsidy – e.g. the increase in units sold plus the government subsidy – is €7,525 (€43 × 175), which means an increase of €3,525.

- *Consumer spending* increases from €4,000 to €6,475 (€37 × 175)

Consumer/supplier surplus and deadweight loss

A subsidy increases both the willingness and ability of suppliers to produce more goods and for consumers to consume more goods – in other words, we can expect both supplier and consumer surplus to increase. However, since the entire cost of the subsidy will be less than the overall increase in societal surplus, there will be a deadweight loss.

Consumer surplus; since the price has fallen from €40 to €37 and quantity consumed has increased from 100 tonnes to 175 tonnes, there is an increase in consumer surplus. Referring to Figure 14.3:

- *Initial* consumer surplus is comprised of areas 1 and 2

  - We need the P-intercept of the demand curve, which 1,100/25 = €44

  - Consumer surplus is calculated as

    $$\frac{(44 - 40) \times 100}{2} = €200$$

- Consumer surplus post-subsidy is $\frac{(44 - 37) \times 175}{2} =$ €612.5 (areas 1, 2, 5 and 6)

- The increase in consumer surplus is therefore €412.5 (areas 5 and 6)

Supplier surplus; suppliers get €43 in total for every tonne produced. That is, €37 comes from consumers and €6 from government.

- *Initial* producer surplus is comprised of areas 5 and 7

  - P-intercept of the supply curve is €36 (see above)

  - Supplier surplus is calculated as

    $$\frac{(40 - 36) \times 100)}{2} = €200$$

- Producer surplus post-subsidy is $\frac{(43 - 36) \times 175)}{2} =$ €612.5 (areas 2, 3, 5 and 7)

- Producer surplus has increased by €412.5 (areas 2 and 3)

Producer revenue

Subsidy = €6 per tonne

S_0

$S_{+subsidy}$

43
40
37

P (€/tonne)

1
2
3
4
5
6
7

D

100   175   Q/t (tonnes/month)

Consumer surplus before subsidy

Consumers surplus after subsidy

Deadweight loss due to subsidy

Supplier surplus after subsidy

Supplier surplus before subsidy
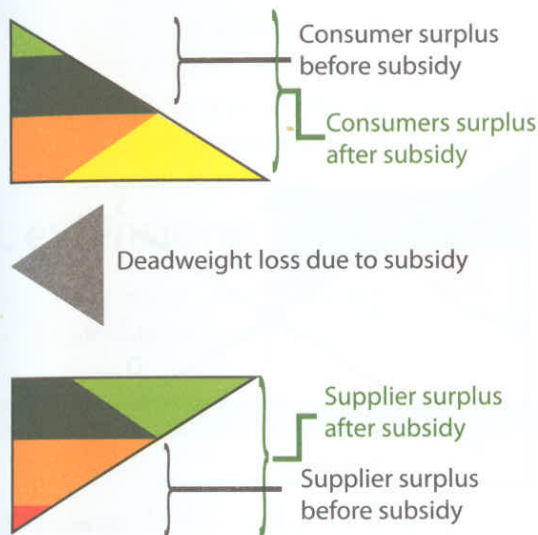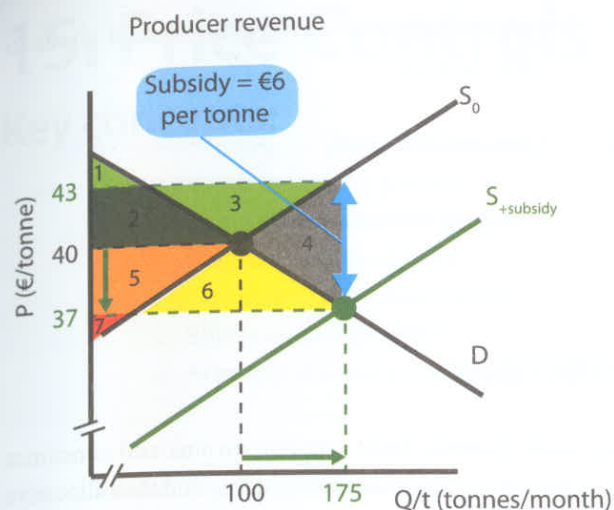
Figure 14.3 Consumer/supplier surplus and deadweight loss

Now for the issue of deadweight loss due to a subsidy – and this is where I lose a few people. Think of a subsidy as the government paying Tom to draw a bucket of water from the well and redistribute it to others in society – in pouring 100 glasses of water, some of the water will drip out. Thus, the total cost to government will be greater than the increase in total utility for those getting a glass of water. Any cost which is not 100% offset by an increase in societal surplus means that there is allocative waste – a deadweight loss.

- Deadweight loss; area 4 in Figure 14.3 shows the deadweight loss.

    o Total cost to government is €1,050

    o The increase in consumer surplus is €412.5

    o The increase in supplier surplus is also €412.5

o The total incidence of the subsidy which does not result in an increase in societal surplus is thus €225 – which is the deadweight loss.

Alternatively, the deadweight loss triangle can be calculated as:

$$\frac{(6 \times 75)}{2} = €225$$

Another way of looking at the deadweight loss is to consider that the marginal costs for producers have been 'artificially' lowered. The de facto marginal cost is given by the supply curve S_0. Marginal utility, or the marginal social benefit, is shown by the demand curve. Every unit produced between 100 and 175 tonnes has a *real* marginal cost that is higher than the marginal benefit. This shows that resources are in fact being misallocated for these 75 tonnes.

POP QUIZ 14.1

1. Assume a supply function of $Qs = -12 + 3P$ and a demand function of $Qd = 24 - 1.5P$.

2. Draw a diagram showing initial equilibrium.

3. Add a subsidy of $4 and calculate the new supply function. Draw a new supply curve showing equilibrium price and quantity after the subsidy.

4. Calculate the total incidence of the subsidy (cost to government) and the incidence on consumers.

5. Calculate the total increase in revenue for producers.

6. Calculate the increase in consumer and supplier surplus.

7. Calculate the deadweight loss.

Summary & revision

c.   Opportunity costs of foregone societal goods in other sectors

d.   Questionable equity

1.   A subsidy is a gift or grant of money to firms which lowers (marginal) costs of production and serves as an incentive for firms to increase output.

e.   Dumping of excess produce on LDCs

2.   Governments subsidise industries in order to create employment, preserve a way of life/culture, create equity in society, improve the home economy's international competitiveness, facilitate R&D, and increase consumption of goods considered to be beneficial to society.

HL extension

5.   Subsidies increase both supplier surplus and consumer surplus but result in a deadweight loss and thus allocative inefficiency

3.   The effects of a subsidy:
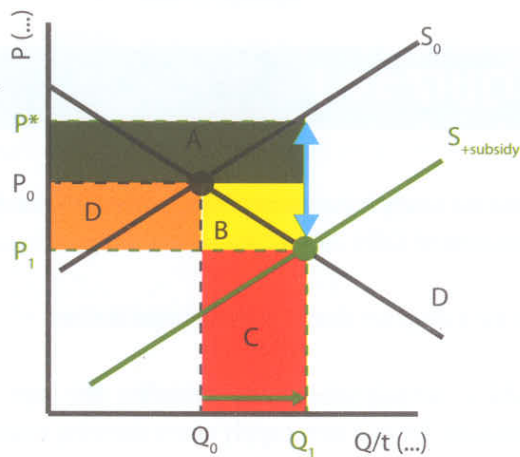
- Lower price for consumers (P_0 to P_1)

- Increased quantity on market (Q_0 to Q_1)

- Increased income for firms (areas A, B and C)

- Cost of subsidy to govt (areas A, B and D)

- Incidence of subsidy on consumers (areas B and D)

- Incidence of subsidy on suppliers (area A)

4.   Subsidies have been seriously criticised for, amongst other things

a.   The hidden costs of administration

b.   Efficiency losses
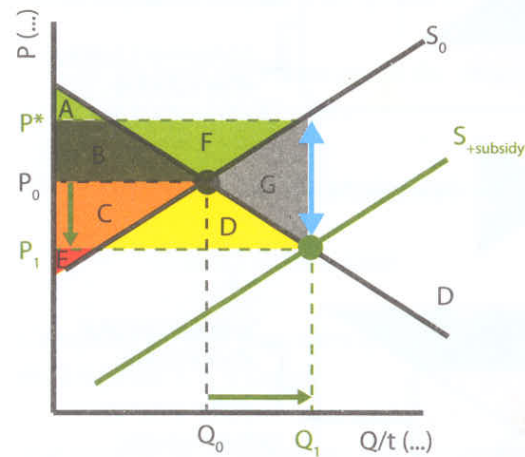
- Consumer surplus before subsidy (areas A and B)

- Consumer surplus after subsidy (areas A, B, C and D)

- Supplier surplus before subsidy (areas E and C)

- Supplier surplus after subsidy (areas E, C, B and F)

- Deadweight loss (area G)

HIGER LEVEL

MICROECONOMICS

15. Price Controls

Key concepts:

- Maximum prices (ceiling prices)
  - o Reasons for maximum prices
  - o Diagrammatical analysis
  - o Outcome of maximum prices
  - o Effects on stakeholders
  - o Attempts at solving market disequilibrium

- Minimum prices (floor prices)
  - o Reasons for minimum prices
  - o Diagrammatical analysis
  - o Outcome of maximum prices
  - o Effects on stakeholders
  - o Attempts at solving market disequilibrium
  - o Minimum wage

HL extensions:

- Calculations...max prices
- Calculations...min prices

Throughout the market iteration thus far, we have operated under the assumption of competitive markets, i.e. markets where only the forces of supply and demand set equilibrium price and quantity. In reality things are far more complex, as there are a number of elements which can offset and negate (= work against) freely operating market forces. Market intervention via controlling prices will have the effect of non-clearing markets with either an excess of supply or an excess demand as a result. We now look at each of these in turn.

Maximum prices (ceiling prices)

A maximum price – also known as a ceiling price – is when the price is set below the market clearing price level. This is a mechanism imposed by government primarily in order to increase availability – perhaps hereby increasing equality in society by permitting more people to afford the good. Notable examples have been war-time centralised prices, extreme measures in times of high inflation and when prices are set centrally in planned economies. In all cases there are two major side-effects. The first effect is an excess of demand which in turn will create queues and second-hand selling, i.e. black (parallel) markets.

Definition: 'Maximum price'
A maximum price – also ceiling price – is a price set down in law that the price may not be set above a certain level. Such intervention often leads to shortages and black markets.

Reasons for maximum prices

The main reason for governments setting a maximum price is for reasons of equity, which means fairness in the distribution of goods. Many countries set maximum prices on goods which are considered basic necessities, such as tortillas in Mexico and cooking oil in Indonesia. The aim of government is to provide broad (as in 'across society and social groups') access and availability of such staple goods for low income groups. The issue with inner city apartments is similar; income groups such as nurses, store clerks and librarians could not possibly afford the market based rents in cities such as Stockholm and New York.

Diagrammatical analysis of maximum prices

Let us assume that a government wishes to set a limit on the rents of inner-city housing – something my home country, Sweden, does in the case of apartments in Stockholm[1]. Since a maximum price only has effect when set below the market clearing price, the result will be an increase in quantity demanded and a decrease in quantity supplied. The free market supply and demand result in the price P_{mkt} and the quantity Q_{mkt} in figure 15.1. (SEK in the figures means "Swedish crowns".)

At a quantity of Q_S m²/year, people are willing to pay $P_{B.M.}$ per m² – far more than the *legal* rent of P_{max}. The maximum market price at Q_S ($P_{B.M.}$) minus the 'initial legal rent' of P_{max} gives the extent of the *possible black (parallel) market*.

As the price is set 'artificially', at below market clearing level (P_{mkt}), the maximum price will result in an *excess of demand* of $Q_S \Leftrightarrow Q_D$.
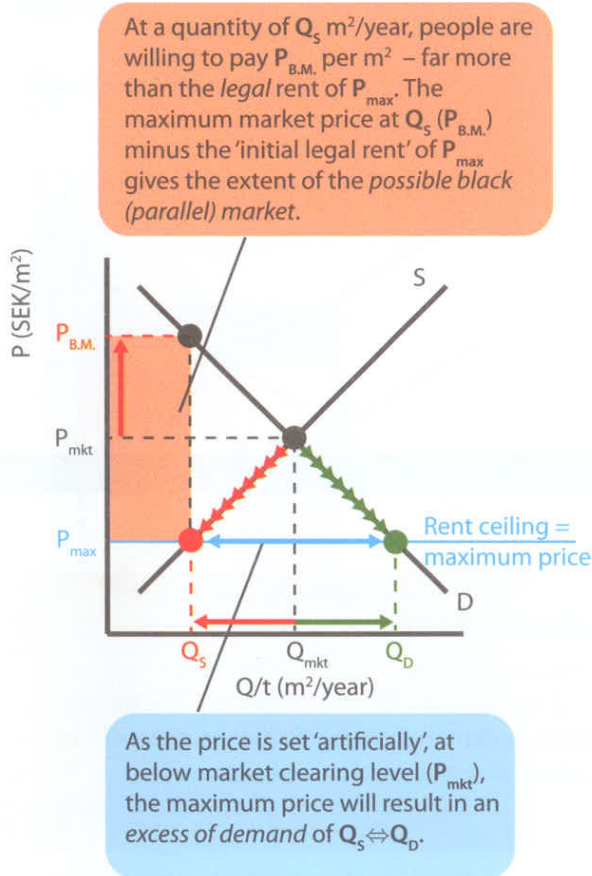
Figure 15.1 Maximum price on inner-city rents, Stockholm

Outcome of maximum prices

"In many cases rent control appears to be the most efficient technique presently known to destroy a city—except for bombing it." (Assar Lindbeck, famous Swedish economist in

1 The queue system works something like this: After giving birth, the happy mother takes the infant down to the official registrar for apartments and puts the toddler in the queue. 20 to 25 years later, the young man/lady might get an apartment. I've had numerous 18 year old students who had been in the queue all their lives. Very patient people, the Swedes. Maybe that's why they drive Volvos.

The Political Economy of the New Left, 1972, page 39)

As quantity demanded is in excess of quantity supplied at P_{max}, (Figure 15.1) there will be a goodly portion of pent-up demand amongst consumers. If no limit is set on the amount of housing one can rent, there is an incentive to rent more apartments than one can use in order to rent out the rest on a parallel (black) market. The black market price – called sub-letting rent – would be at a price of up to $P_{B.M.}$ since this is what consumers would be willing to pay for the quantity of Q_S according to the demand curve. The red quadrant between the minimum reselling price (the official maximum price of Pmax) and the black market price ($P_{B.M.}$) is the possible black market. Note that the black-market price of $P_{B.M.}$ is based on the assumption that all of Q_S housing hits the parallel market! Hence the word "possible" above.

The consequences of imposing maximum prices are the queues and resulting black markets. This is something that the government will have to deal with and the most common form of solution has been to limit the quantity per person by either a rationing system, or a queuing system. Rationing is achieved by setting a limit to purchases and such instruments as coupons for coffee and meat, while a queue system is basically done by instituting a 'first come – first served' system often found on markets for rent-controlled inner-city apartments. Note that it depends very much on the good in question which of these are possible for government to impose – a rationing system works better for coffee than for housing.

Effect of maximum prices on stakeholders

Since the price of the good – here, inner-city apartments – decreases one would expect consumers to be better off and suppliers to be worse off. This is, well, almost true. Figure 15.2 shows how the maximum price (rent ceiling) results in a redistribution and loss of societal surplus.

- Consumer surplus is originally areas A, B and E. The maximum price means a loss of area E but a gain of area C.

- Supplier surplus is originally areas C, D and F. Due to a maximum price, suppliers lose areas C and F.

- Consumers can be considered better off since the maximum price transfers a portion of supplier surplus (area C) to them which offsets the loss of consumer surplus (area E).

Suppliers are worse off since they lose supplier surplus areas C and F.

The net loss of consumer surplus and supplier surplus (e.g. societal surplus) are areas E and F. This shows the efficiency loss of the maximum price – the deadweight loss.
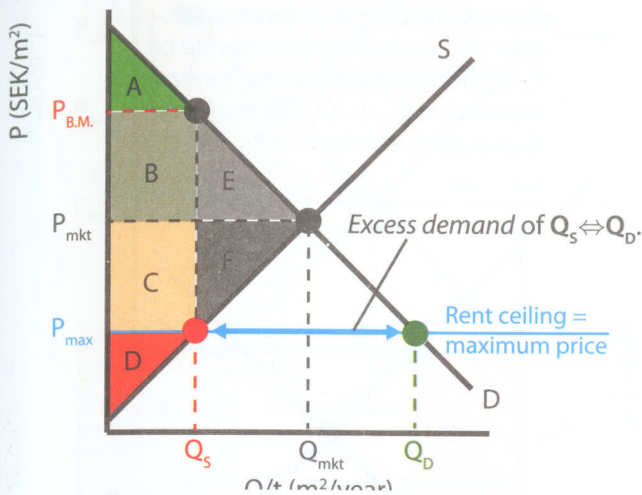
Original consumer surplus is A + B + E and after the rent ceiling is imposed becomes A + B + C. The loss of consumer surplus (E) is offset by a gain of C which has been transferred from suppliers.

Original supplier surplus is C + D + F. After the rent ceiling is imposed, remaining supplier surplus is D.

Area E is the net loss of consumer surplus and F is the net loss of supplier surplus. This is the total loss of societal surplus due to the maximum price. This shows suboptimal efficiency – a deadweight loss.

Figure 15.2 Consumer and supplier surplus due to a maximum price (rent ceiling)

There are some additional inefficiencies and re-distribution effects as a result of maximum prices:

- The lucky ones who manage to get hold of the good (apartments) will be able to re-sell or sub-let the apartment and earn profits at the expense of others[2]. Areas A and B in figure 15.2 also represent the

potential black market for apartments (see Figure 15.1). Any apartments rent out at a price of $P_{B.M.}$ means that those units would have a corresponding decrease in consumer surplus. Just imagine if all the apartments were sold on the black market – areas B and C would be profits for the renters and remaining consumer surplus would be area A.

- Rent ceilings would create a disincentive for apartment owners to keep the apartments on the market (see 'A little depth: rent controls…' further on) and they would look for producer substitutes such as renting out to businesses or storage. This would decrease the supply of apartments and raise the black market price further.

- Another effect would be that apartment owners would provide minimum possible upkeep, renovation and repairs. This would destroy capital over time and possibly result in 'slummification' in areas with rent control[3].

- For inner-city rents it is probable that maximum prices cause immobility on the labour market as it might be quite difficult to get the same low rents in other cities. In essence, a person finding a job in another city might not be willing to move.

New York - Capital City of the Apartment Lifestyle

---

2   In fact, 'lucky ones' is not the right term. The correct term would be 'those with lower opportunity costs of standing in line'. Just think about who is selling black market tickets and who is buying them! Correct: my kids stand in line for me and then sell me the Rammstein tickets at a markup. They have more time than money – I have the reverse, so my opportunity costs are greater than theirs. (In fact, my kids are so wonderful that they give the tickets to me as a present (!) but I'm trying to make a point.)

3   I recommend "Economic Facts and Fallacies" by Thomas Sowell, pages 23 – 36 for an excellent romp through what is in effect a case study of unintended consequences.

A Little Depth

Rent controls on inner city apartments – long run effects

One thing I make sure my people know is that I am training them not to by my equal but to be my superior, and I always succeed. I encourage them to go off to university after graduation and return to straighten me out. Many of them do. One young lady, Sara, came for a visit to discuss how price elasticities can explain the limited supply of inner-city housing. This is HER story.

"When a maximum (ceiling) price is put on desirable inner-city housing, there will be an excess of market demand, Q_1 to Q_2 in the diagram", explained Sara. "Right?"

"Yupp", I said, "but what's with the new supply curve?"

"Well, it is foreseeable that landowners over time simply don't feel that the rent is worth it, and would take rooms and areas off the housing market."

"For what?! I mean, better to at least get something for it!"

"Ah, but you are forgetting the opportunity costs! These rooms, attics and other space might be better used for other things; office space, ateliers, gyms or simple storage. The maximum price governs living area, not other activities!"

"Hmmm, OK, I'm with you on the 'shift' part, but what's the "swivel" and thus increase in PES?"

"Simple", said Sara with that special smile young women reserve for old men who aren't quite with it. "As the owners have converted some of the available floor area into other uses, there are now an increased amount of producer substitutes available! Should the price ceiling be lifted, suppliers will be able to quickly convert floor space back into living areas for rent. We assume that demand is the same, which means that full-out supply of rooms would be as before; at P_{mkt} and Q_{mkt}."

I thought this was a pretty cool use of PES. Thanks Sara. Incidentally, all of you may feel perfectly free to excel beyond my level and thereupon come and educate me! Welcome.

> In the long run (LR), owners of apartments and housing space will convert free space into other uses – e.g. producer substitutes. This decreases supply at all levels up to what the free market price would be, P_0. PES has increased. Excess demand increases to $Q_{LR} \Leftrightarrow Q_2$ and the black market price increases to $P_{B.M.2}$.
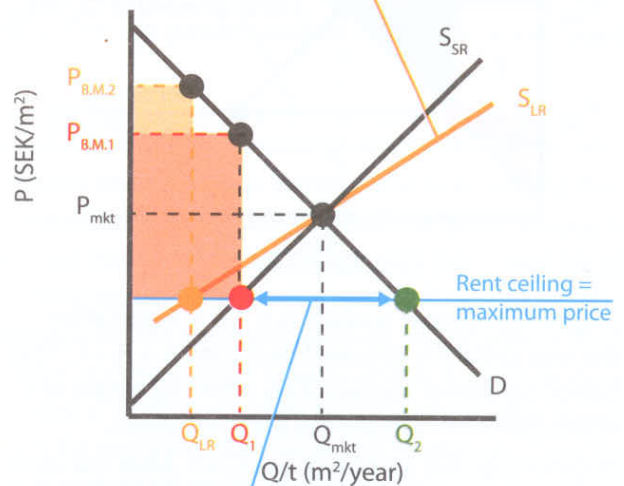
> The ceiling price P_{max} creates an initial excess demand for housing of Q_1 to Q_2.

Figure 15.3 Effect of rent controls in LR

Attempts at solving market disequilibrium in maximum price situation

There is also the possibility of using market forces to move back to equilibrium. Let us look at two general possibilities; shifting supply or demand for inner-city Stockholm housing where a maximum price has been set. In Figures 15.4a and b, the maximum price creates an excess demand of $Q_S \Leftrightarrow Q_D$.

Figure a) Increasing supply

A *subsidy* and/or government supply of new housing increases supply, removing the excess demand of $Q_S \Leftrightarrow Q_D$.
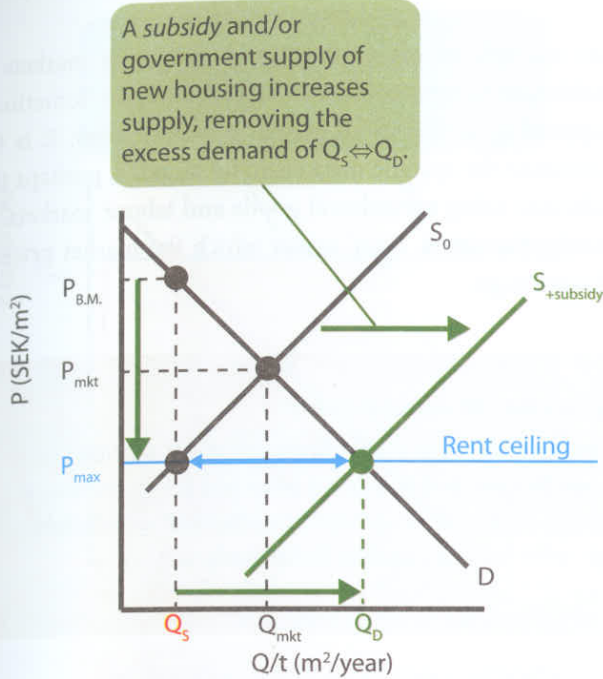
Figure b) Decreasing demand

Government inducements for *substitutes* to inner-city apartments lowers demand, removing excess quantity demanded.
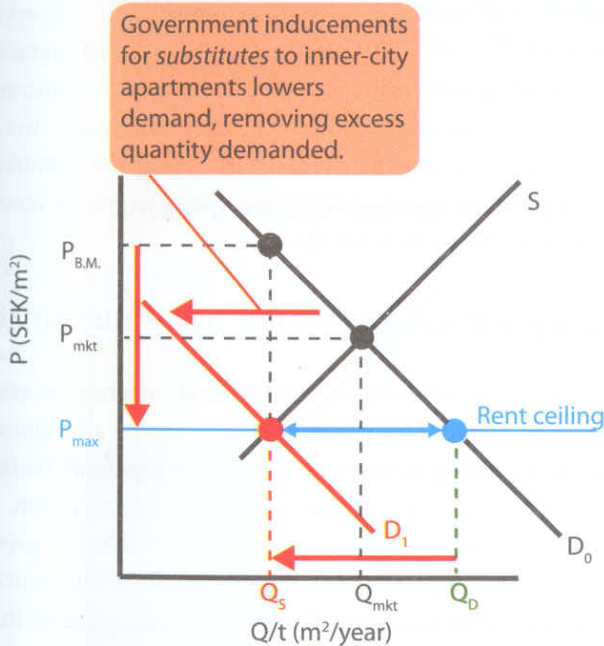
Figure 15.5 Possible solutions to the housing shortage

The government could *subsidise* cheap inner-city housing by offering low cost loans to building companies or by offering incentives for city councils to increase the amount of apartments. This would shift the supply curve from S_0 to $S_{+subsidy}$ in Figure 15.4a), which would do away with the shortage of housing – and the black market.

Alternatively, government could offer any number of incentives for people to forego inner-city living by enhancing the alternative (= outer city areas), which would lower demand for city housing. For example, increased/improved transportation to outer city areas, tax benefits for those who commute, or even by increasing certain inner-city specific taxes – say parking fees and traffic fees – which are all ways for government to change citizens' living preferences. This is illustrated in Figure 15.4b), where the decrease in demand (D_0 to D_1) lowers the black market price to the official maximum price, thereby obliterating the excess demand for inner-city housing.

Putting the pieces together

Born again in the USA.

No, the title is not a reference to the Republican vice-president candidate in USA 2008, but to one of Bruce Springsteen's most famous songs.

Bruce Springsteen came out with a new album in 2002 – the first with his original E-Street Band since 1987. Basically, the album was in commemoration of the September 11th attacks in New York in 2001 and came out on the anniversary. The hype was of such magnitude that Springsteen's 2003 tour was virtually sold out in a matter of hours. In Göteborg (Gothenburg) Sweden (on the 4th of November 2002), all 50,000 tickets for the concert on the 21st of June 2003 were sold within 40 minutes and the queues snaked around the entire city block! This was quickly resolved by organising another Springsteen concert for the 22nd of June.

This was, of course, not entirely due to Springsteen's undying devotion to his fans. Two reasons were put forward by the organisers. The first was that one wished to avoid the chaos of scalpers (= black market ticket sellers) and black market tickets by trying to accommodate demand. Secondly, more tickets simply meant more revenue! 'Bums on seats' and all that.

The potential black market (or parallel market) is between P_0 and $P_{B.M.}$

By setting the original price (P_0) below the market equilibrium ($P_{B.M.}$), a ceiling (maximum price) was created, causing an excess of demand (A⇔B), i.e. a shortage of 50,000 tickets.

Initial excess demand

$Q_{tickets}$ 21st & 22nd June 2003 (1,000s)
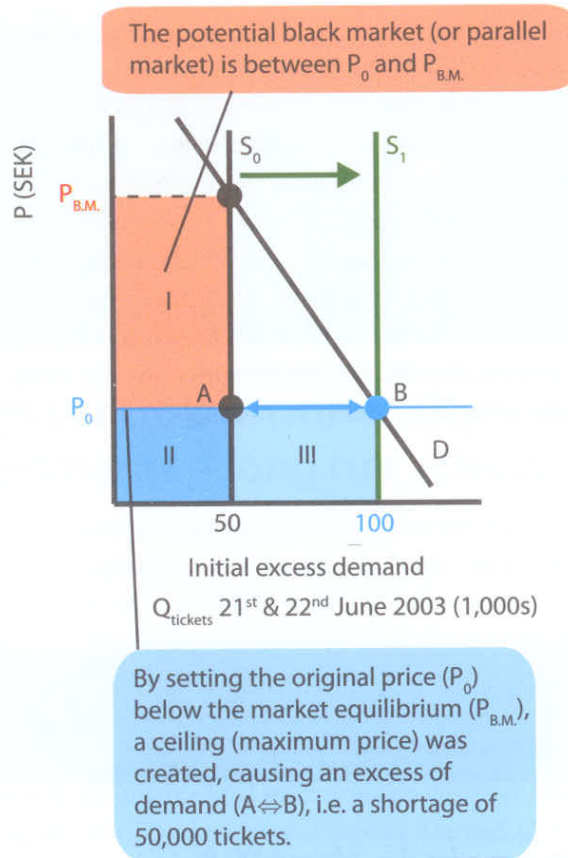
Figure 15.5 Tickets for Springsteen concert

This scenario is shown in Figure 15.5. The initial supply curve is S_0 and is perfectly vertical (= demand is completely price inelastic in correct economic jargon) at 50,000 seats. At that price, there would be a possible black market created by the excess demand (A⇔B) and the high demand is an incentive for re-selling tickets on the 'second hand' market. This is shown by the shaded area I, which is given by the intersection of the original supply curve and the demand curve. One can say that anyone fortunate to get hold of a (first hand) ticket would be able to resell it at the black market price of $P_{B.M.}$. (Note that this is the possible black market price and not necessarily an equilibrium price.)

By holding an additional concert, the organisers have increased the supply of tickets from S_0 to S_1. This should hopefully result in the market equilibrium price of P_0 and destroy any black market in the making. One can also see how the total revenue (price times quantity) increases from area II to area II + III. One might say that this move has swept the market out from under the ticket scalpers' feet.

Minimum prices (floor prices)

Another example of government intervention on markets is the establishing of minimum prices. While this has sometimes been applied generally to all goods in an economy, it is far more common for specific markets to be targeted, perhaps the most obvious being agricultural goods and labour markets. A minimum price sets a 'floor' under which the market price is not allowed to go.

Definition: 'Minimum price'
A minimum price – also floor price – is a price set down in law that the price may not be set below a certain level. Often governments must guarantee the price by purchasing the excess supply.

Reasons for minimum prices

The intention of minimum price is to protect and aid certain suppliers; a minimum price on agricultural goods will guarantee farmers what government considers an acceptable income, while minimum prices on labour – i.e. minimum wage – would benefit those supplying their labour. In setting the minimum price on a good, the government is attempting to benefit society – an outcome that is often not the case.

Diagrammatical analysis of minimum prices

Government intervention on agricultural markets is often motivated by wanting to preserve a landscape or a traditional way of life by aiding farmers in keeping an equitable (= fair) standard of living, e.g. similar to that in other sectors in society. By setting minimum prices on agricultural goods, governments can even-out income differentials (= differences) by guaranteeing that farmers will receive a certain price for their goods. In doing this, the government puts the market function out of order – in essence by guaranteeing that farmers will get a minimum 'fair' price for their produce.

a) Costs of repurchasing

As the price is set "artificially", at P_{min} – above market clearing level (P_{mkt}) – the minimum price will result in an *excess supply* of $Q_D \Leftrightarrow Q_S$.
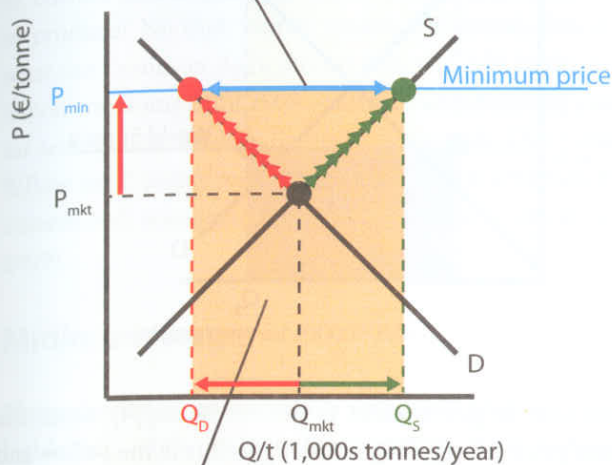
As the government supervisory authority cannot allow the excess to hit the market, the excess must be removed from the market. The total cost to government of this repurchasing scheme is $P_{min} \times Q_D \Leftrightarrow Q_S$, which is shaded yellow area in the diagram.

b) Increase in revenue

Increase in producers' revenue
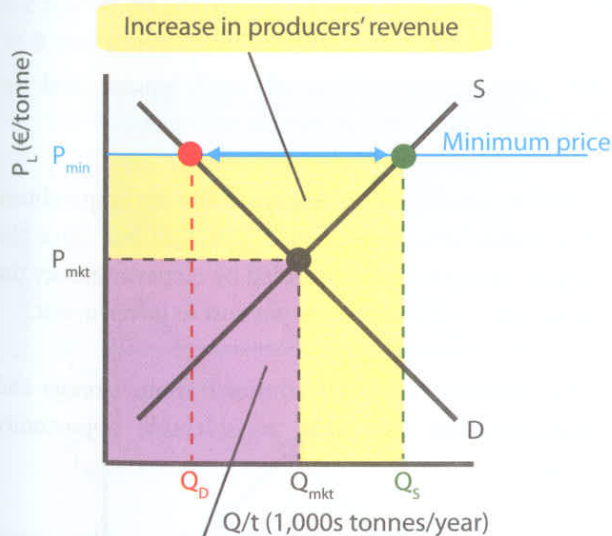
Producers' revenue before minimum price scheme.

Figure 15.6 Minimum price on grain – costs and revenue

A price support scheme simply means that the government agrees to purchase the excess at the agreed minimum price. In Figure 15.6a), the price rises from P_{mkt} to P_{min} and quantity demanded decreases from Q_{mkt} to Q_D. The total excess amount of grain is $Q_D \Leftrightarrow Q_S$, which the government would have to buy at a price of P_{min}. Total government expenditure is thus the orange quadrant. Producers' total revenue, Figure 15.6b, increases by the yellow 'boomerang shaped' area. One should mention that other costs linked to the minimum price scheme will arise, such as administrative costs, storage costs, and transportation costs – an estimated 60% of the total cost of the European Union's (EU) Common Agricultural Policy (CAP) paid by taxpayers went to storage and administrative costs[4]. Anyhow, the government portrayed in our example now has a few hundred thousand tonnes of grain to deal with. Now what?

Outcome of minimum prices

Some of the most wasteful acts in society, tragically enough, often agricultural surpluses have simply been stored in warehouses, resulting in 'grain heaps', 'beef mountains' and 'wine lakes' which nobody seems quite sure what to do with. In many cases the surplus has been burnt, sold on other markets (see 'dumping'), or even sold back to farmers at a fraction of the minimum price – which was then often used as cattle feed to produce more butter and beef… I have to be careful here. My students tell me I have a tendency to get very loud and froth at the mouth when I get to government involvement in agricultural output[5]. Suffice it to say that many of the minimum price schemes used in agricultural policies have historically been very wasteful, since suppliers have often produced too much, consumers have paid unnecessarily high prices and developing countries have seen their markets disrupted by excess produce dumped on their countries. We will return to the highly inflammatory debate on agricultural policies in Chapters 65 and 84.

Effects of minimum prices on stakeholders

The price rises and the consumer pays a higher price – obviously consumers are getting a bad deal. How bad? If we follow Figure 15.7 we again will see both redistribution issues and allocative losses:

- Consumer surplus is originally areas A, B and D. The minimum price means a loss of areas B and D.

---

4 Wall Street Journal, 'A rotten harvest', by Richard Howarth, June 29th 2000.

5 My IB2s recently banned me from using any type of energy drink during school hours – and received strong support from colleagues in adjacent (= nearby) rooms. I fear coffee and sugar are next. That leaves only my tequila and whiskey shelf. Oh well.

- Supplier surplus is originally areas C and E. Due to a minimum price, suppliers gain area B.

- Consumers can be considered worse off since the minimum price transfers a portion of consumer surplus (area B) to suppliers.

- Suppliers are better off since the loss of supplier surplus (area E) is more than offset by the increase in supplier surplus (B).

- The net loss of consumer surplus and supplier surplus (e.g. societal surplus) are areas D and E. This shows the efficiency loss of the minimum price – the deadweight loss.

The total cost to government is the excess supply times the minimum price, i.e. $Q_D \Leftrightarrow Q_S$ times P_{min}. This is the yellow and purple area. If the government manages to sell this excess supply on the world market, the revenue to government will be $Q_D \Leftrightarrow Q_S$ times P_{world} – the purple area. The net cost to government is therefore the yellow area. Naturally, the area(-s) representing the cost to government also represent the costs to tax payers.

Figure 15.7 b) Min price on grain – government, tax-payers and LDCs

- **Consumer surplus** before minimum price; areas A, B and D

- **Consumer surplus** after minimum price; area A

- **Supplier surplus** before minimum price; areas C and E

- **Supplier surplus** after minimum price; areas C and B

- **Deadweight loss**; D and E

Figure 15.7a) Min price on grain – government, tax-payers and LDCs

There are, once again, additional inefficiencies and re-distribution effects as a result of minimum prices:

- The consumers are also taxpayers! The total expenditure for consumers is the area $P_{min} \times QD$ but since the repurchasing scheme is funded by taxpayer money the total cost must include the net cost to government.

- The additional costs of administration, storage and transportation can mean considerable opportunity costs.

- If the government or suppliers dump the excess grain on the world market the result is often that the world price decreases. This can have severely harmful effects on developing countries which are dependent on exports of primary goods. Basically the effect is to redistribute income from poor people in poor countries to rich people in rich countries.

MICROECONOMICS

Attempts at solving market disequilibrium in minimum price situations

Apart from buying up the excess and destroying it or selling it at below production costs, what can governments actually do to correct this oversupply? How about…paying farmers not to produce? No that sounds insane. Of course, this is exactly what the Common Agricultural Policy (CAP) in the European Union was doing until 2009 – farmers were basically paid a sum for setting aside fields to decrease oversupply. This 'set-aside' or 'fallow-field' policy was subject to so much criticism from EU citizens and non-government agencies that it was abolished in 2009[6].

Minimum wage

Many countries' governments impose a minimum wage via legislation, while many other countries have a de facto (= actual) minimum wage through the influence of strong unions and centralised wage agreements with employer organisations. The reasoning behind minimum wages is that weaker members of society (e.g. workforce) need support; unskilled labourers, young people entering the labour market, people with little job experience, minority groups, etc. There is also, once again, an element of attempting to even-out income differentials in society. Whatever the case, the market for labour is similar to the market for goods and services, which can be seen in Figure 15.8, and shows the effects of a minimum wage.

The minimum wage is set at $P_{L_{min}}$, above the market clearing price. This results in a surplus of labour at the going wage rate, i.e. increased unemployment of $Q_{L_D} \Leftrightarrow Q_{L_S}$.
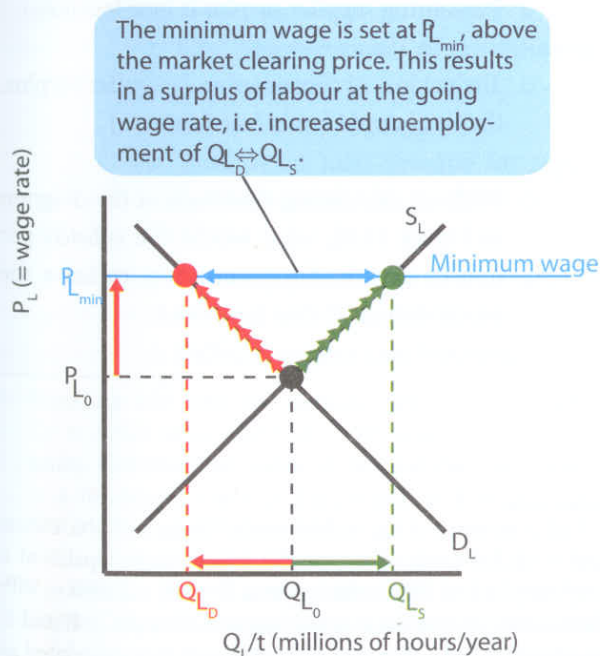
Figure 15.8 Minimum wage

6 See the European commission at http://ec.europa.eu/agriculture/healthcheck/index_en.htm

The **supply of labour** (S_L) shows the propensity of the labour force to accept jobs at given wage rates, while the **demand for labour** (D_L) shows the willingness of firms to hire at given wage rates. Equilibrium wage on a perfectly free market would be P_{L_0} and the amount of people employed would be Q_{L_0}. A minimum wage above the market equilibrium means that more people offer themselves on the labour market, shown by the increase from Q_{L_0} to Q_{L_S}. At the set minimum wage of $P_{L_{min}}$, firms' demand for labour decreases from Q_{L_0} to Q_{L_D}. This strongly resembles the outcome in the previous example; there is a surplus of labour, otherwise known as an increase in unemployment, shown in the diagram as the distance from Q_{L_D} to Q_{L_S}.

QUESTIONS:

1. What would happen to the wage rate if the demand for labour increased so that the demand curve intersected the supply curve somewhere below $P_{L_{min}}$?

2. Same scenario as in question one; what would happen to the level of unemployment?

3. What would happen on the labour market if the minimum wage rate were lowered?

4. What would happen on the labour market if demand for goods and services in the economy increased? (Hint; 'derived demand'.)

The effects of minimum wages are the subject of a very hot political debate. Defendants of minimum wages argue that many labourers would otherwise be powerless on the labour market since firms could set wages at close to existence minimum for weak labour groups. Opponents argue that minimum wages add to unemployment and lead to inefficient labour markets, resulting in sub-optimal resource allocation. We will look into this in greater depth in Chapter 50.

Exam tip; using the minimum price diagram in analysis

The basic order of progression in analysing the effects of a minimum price is the following (refer to Figures 15.6 and 15.7):

1. Initial outcome: excess supply ($Q_D \Leftrightarrow Q_S$).

2. Initial possible government response: repurchasing scheme to get rid of excess supply.

3. Winners and losers I: higher prices for consumers but increased revenue for suppliers. The cost of the repurchasing and storage all come from taxpayer monies.

4. Secondary possible government responses: the excess must be dealt with, e.g. via storage, destruction or dumping.

5. Long run scenarios (winners and losers II): the government earns revenue by dumping on other markets which can recover some of the costs of repurchasing and storage. If the excess is dumped on a developing country which cannot compete with the dumping price, markets are destroyed and livelihoods lost.

POP QUIZ 15.1

MAXIMUM AND MINIMUM PRICES

1. Using diagrams, explain what happens in terms of optimal resource allocation when a minimum or maximum price is put on a good.

2. Explain why a government cannot put a maximum price on a good without other measures. Illustrate your answer with an appropriate diagram.

3. In the same diagram as in question 2, show total government costs of the minimum price scheme. Are these the total costs to government of the scheme?

4. A government institutes a minimum price scheme for agricultural goods. What are the likely effects on farmers' incomes? Use a diagram and show total income for farmers.

HL extensions

Calculating the effects of maximum and minimum prices is very similar to what you have already done in previous calculations – here it is basically a matter of having a firm grip on the supply and demand functions and then popping in a different price than that given by market equilibrium.

Calculations; maximum prices

The demand function is $Qd = 2,000 – 50P$ and the supply function is $Qs = 0 + 50P$. Figure 15.9 shows market equilibrium established at a price of $20 and 1,000 units.

- Governments sets a maximum price of $8:

    o Inserting $8 into the *demand* function gives (see footnote[7])

    o *Quantity supplied* will be (see footnote)

    o *Excess* demand will be (see footnote)

    o At a quantity of **xx**, a potential *black market* price is established at $P_{B.M.}$ (see footnote)

    o Remaining *consumer surplus* is (see footnote)

    o Remaining *supplier surplus* is (see footnote)

    o The net loss of consumer and supplier surplus, the *deadweight loss* is (see footnote)

    o Without calculating (just look at the diagram in Figure 15.9), what would the subsidy per unit have to be for government to halve the excess demand? (See footnote.)

---

7 $Qd = 1,600$ units and $Qs = 400$ units. There will be excess demand of 1,200 units. The possible black market price at a quantity supplied of 400 units is given by $400 = 2,000 – 50P$; $32. Remaining consumer surplus (areas A, B and C): B and C are calculated as $(32 – 8) \times 400 = \$9,600$; area A is calculated as $([40 – 32] \times 400)/2 = \$1,600$. Remaining supplier surplus is $(8 \times 400)/2 = \$1,600$. Deadweight loss is $([32 – 8] \times 600)/2 = \$7,200$. The subsidy would have to shift the supply curve down by $12 in order for the supply curve to intersect the minimum price at 1,000 units.
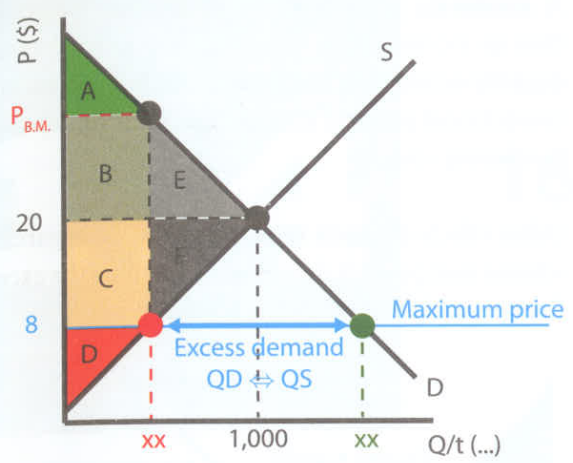
Figure 15.9 Maximum price

Calculations; minimum prices

By now I suspect that the calculations are becoming almost 'mechanistic' for you, so I shall use the same supply and demand functions as above.

- A minimum price of $26 gives the following outcomes:

  o Quantity supplied and quantity demanded (see footnote[8])

  o Excess supply (see footnote)

  o Cost to government of repurchasing scheme (see footnote)

  o Increase in total revenue for suppliers (see footnot)

  o Remaining consumer surplus (see footnote)

  o Supplier surplus (see footnote)

---

8 Inserting $26 into the demand and supply functions gives Qs = 1,300 units and Qd = 700 units. Excess supply is 600 units. Government needs to remove the 700 units of excess at a minimum price of $26, which is $18,200. Initial TR is $20,000 and after the minimum price is imposed $33,800 – the difference is $13,800. Consumer surplus after the imposition of a minimum price is area A, calculated as ([40 – 26] × 700)/2 = $4,900. Supplier surplus (areas B, C and D); areas B and C are calculated as [26 – 14] × 700 = $8,400; area D is (14 × 700) / 2 = $4,900; total supplier surplus is thus $13,300. Deadweight loss is given by ([26 – 14] × 300)/2 = $1,800. If the government sells the excess at $4 per unit, the revenue is $4 × 600 = $2,400 and this offsets a portion of the original buy-back of $18,200; the net cost is therefore $15,800.

---

o Net loss of consumer and supplier surplus, deadweight loss (see footnote)

o The government sells the excess on the world market for $4 per unit. What is the *net* cost to government – disregarding administration, storage, transportation and opportunity costs? (See footnote.)
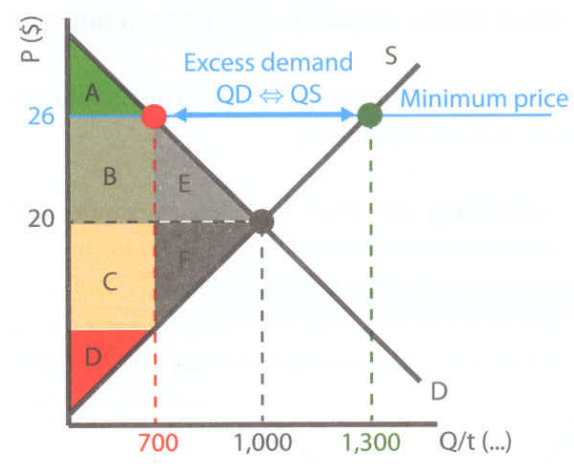
Figure 15.10 Minimum price

Summary & revision

1. A maximum price (also 'ceiling price') is a government set price above market equilibrium price. The intention is to increase wider availability of the good.

2. A maximum price leads to a decrease in price and thus an increase in quantity demanded and a decrease in quantity supplied. The initial result is an excess demand.

3. Governments can attempt to countermand the excess demand via a queuing or rationing system.
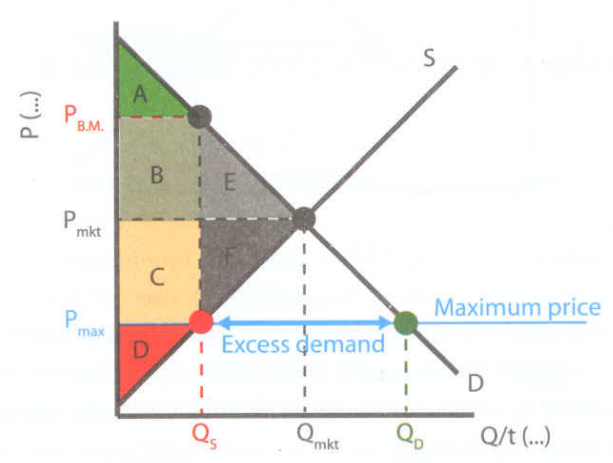
- Lower price for consumers (P_{mkt} to P_{max})
- Decreased quantity on market (Q_{mkt} to Q_S)
- Excess demand ($Q_S \Leftrightarrow Q_D$)
- Possible black market (A and B)
- Increase in consumer surplus (C)
- Decrease in consumer surplus (E)
- Decrease in supplier surplus (C and F)
- Remaining supplier surplus (D)
- Net loss of societal surplus (E and F) is the deadweight loss

4. Governments can attempt to solve market disequilibrium caused by maximum prices by

   a. Subsidising the good (to increase supply and decrease excess demand)

   b. Lowering the price of a substitute, for example via subsidies (to decrease demand and excess supply)
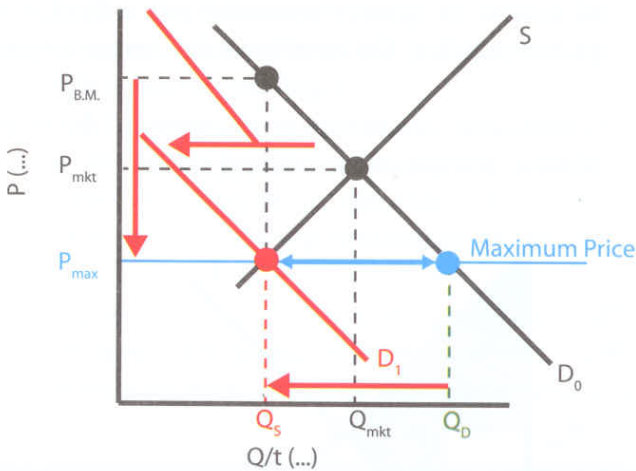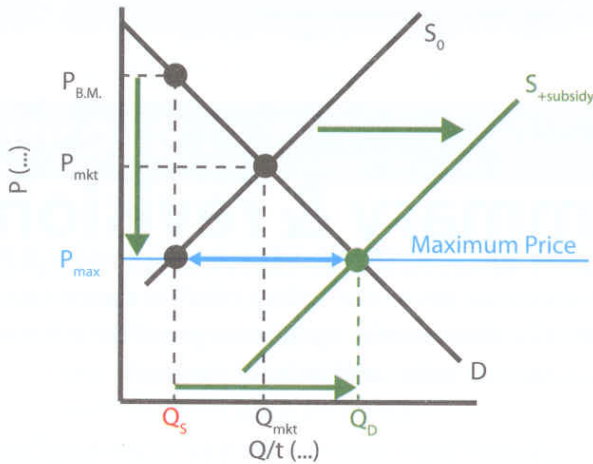
5. A **minimum price** (also 'floor price') is a government set price above market equilibrium price. The intention is to provide certain societal groups with an equitable standard of living.

6. A **minimum price** leads to an increase in price and thus an increase in quantity supplied and a decrease in quantity demanded. The initial result is an excess supply. Government needs to remove the excess supply via a re-*purchasing scheme*.

7. Other effects are **costs to taxpayers** of the repurchasing scheme and possible dumping – i.e. selling the excess on foreign markets at below the costs of production.
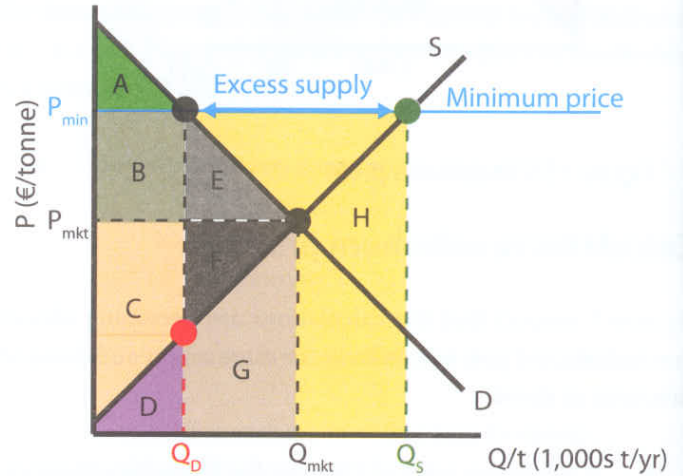
Higher price for consumers (P_{mkt} to P_{min}). Increased quantity supplied on market (Q_{mkt} to Q_S). Excess demand ($Q_D \Leftrightarrow Q_S$). Decrease in consumer surplus (B and E). Remaining consumer surplus (A). Increase in supplier surplus (B). Decrease in supplier surplus (F). Net loss of societal surplus (E and F) is the deadweight loss. Cost of minimum price scheme (E, F, G, and H). Increase in suppliers' revenue (B, E, and H)

8. Another common form of minimum price is **minimum wage**. This is set above the market equilibrium price for labour in order to guarantee low-income groups a minimum income. This can in fact result in a *higher unemployment rate*.

16. Market Failure

1.4
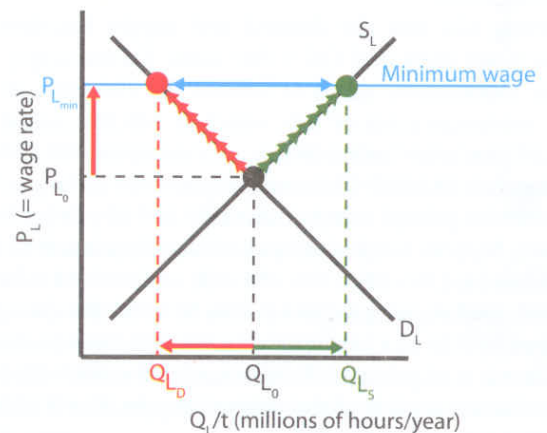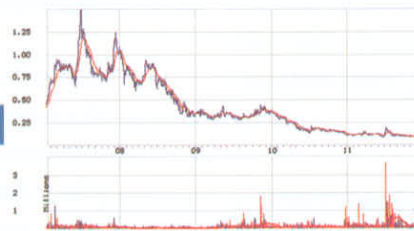
Key concepts:

- Concept of market failure
    - Marginal *private* costs and benefits
    - Marginal *social* costs and benefits
- Types of market failure
    - Externalities
    - Under- and overprovision of goods
    - Imperfect competition
    - Common access resources
    - Information asymmetry (HL)

We have expounded on how markets succeed in both the production and distribution of goods – and how the price function ultimately creates a market clearing. Recall that when quantity supplied equals quantity demanded, the market is in equilibrium and there is optimal resource allocation – a Pareto-optimum. Should – for some reason – the market not achieve this equilibrium then the market has failed.

Concept of market failure

Our premise in dealing with a free and competitive market is that the three main actors in the market – firms, consumers and owners of factors of production – will act in their own self-interest, i.e. all three groups will attempt to optimise their returns. Firms will maximise profits; consumers their utility and factor owners their return. In accordance with standard

economic theory, allowing a 'free' market to reign will produce the best possible outcome in terms of resource allocation. This is the Smithian world of the 'invisible hand'; billions of individual transactions would together create a system where the **right** goods were produced in the **right** amounts. It is in defining 'right' that the concept of market failure arises, specifically when resources are not used in the best possible option.

> **Definition: 'Market failure'**
> When the price mechanism fails to allocate resources towards the socially optimal amount, i.e. there is suboptimal resource allocation, one speaks of **market failure**; there is a non-Pareto outcome.

Marginal *private* costs and marginal *social* costs

If one buys a good on a competitive market then theoretically one is paying what market equilibrium dictates, e.g. supply equals demand and that is the market price. The market price will equal the cost of the last unit produced and both parties are happy with the transaction. However, if there are additional costs of producing and consuming the good but these are instead inflicted upon *third parties* (others in society not involved in the transaction) then there are social costs not accounted for in the private costs of production.

At any quantity where the **MSC > MPC** there is a *negative* externality

MSC

MPC (=S)

Q₀ Q₁ Q₂

Q/t (...)
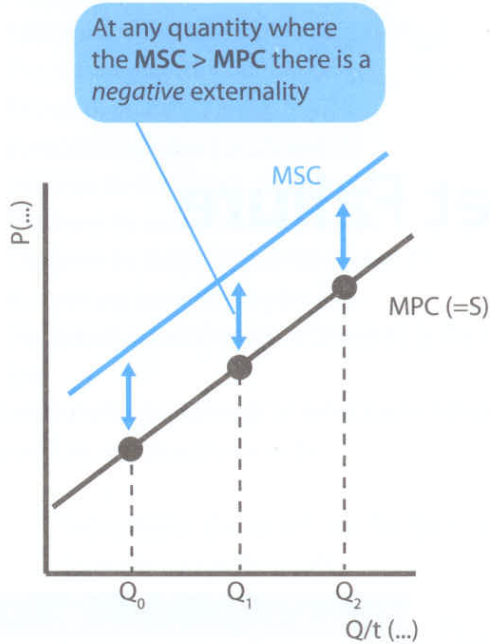
Figure 16.1 Marginal private and social costs (negative externalities)

- The upward sloping *MPC curve* in Figure 16.1 shows how firms' costs rise as output rises. (HL will know this from theory of the firm.) The marginal private cost is basically the sum of individual firms' marginal costs – the market *supply curve*.

- If there are costs inflicted on third parties in producing or consuming the good – for example pollution from using petrol – then one speaks of *negative externalities* ('negative spill-over effects'). The double-edged arrows show the extent of the negative externalities at any given output. **MPC + negative externalities = MSC**.

- The *marginal social cost* of a good shows the marginal cost of society as a whole in providing a good. The **MSC curve** is the *supply curve* (**MPC**) plus negative externalities (pollution).

It is worth noting that we are assuming that the additional costs to society – pollution – can be estimated and put into money terms. This is easier said than done as we shall see.

Marginal *private* benefits and marginal *social* benefits

In purchasing any given good one is trying to maximise ones personal benefit – *marginal utility* in economics. With price as a signal, consumers buy more of the good at lower prices and

thereby increase their total utility. The addition to total utility is the *marginal private benefit* derived from consuming the good. If this consumption has positive effects on third parties then there are social benefits not accounted for in consumers' private benefits. In addressing the issues of positive externalities, we use two terms; marginal *private* benefits (MPB) and marginal *social* benefits (MSB).

- The downward sloping *MPB curve* in Figure 16.2 shows how the marginal private benefits to the private individual (the consumer) fall with increased consumption. This is of course the *marginal utility* previously shown by the demand curve.

- When the marginal social benefit (MSB) is greater than the marginal private benefit (MPB) there are *positive externalities* (or 'spill-over effects') – e.g. positive effects for third parties. This is shown by the double-edged arrows between the MSB and MPB curves in the left diagram; **MPB + positive externality = MSB**.

- The *marginal social benefit* of a good shows the marginal utility of society as a whole in consuming a good. The **MSB curve** is the *demand curve* (**MPB**) plus positive externalities.

At any quantity where the **MSB > MPB** there is a *positive* externality

MSB

MPB (=D)

Q₀ Q₁ Q₂

Q/t
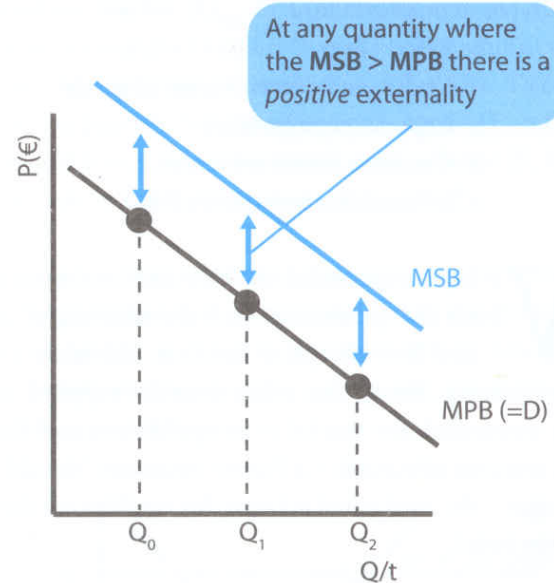
Figure 16.2 Marginal private and social benefits (positive externalities)

Types of market failure

Chapters 17 to 21 deal with the main types of market failure. Keep in mind that the over-riding issue here is that the competitive market has failed to produce the right goods in the right quantities at the right price.

Externalities

When the steel mill in my home town in Sweden, Åkers International, led spill water from the plant into Lake Visnaren for some 500 years, the effects became all to obvious in the 1970s. Increasingly polluted water via sulphates was causing the lake to grow over with weeds and fishing stocks were decreasing. This is a clear example of when external costs to third parties (fisherman and lake house guests) exceed the costs to the firm in producing canons and industrial rolls and the price paid by the firm's customers. The marginal social cost was greater than the marginal private cost – a *negative externality* thus. Pollution is one of the main examples of negative externalities. More on this example in Chapter 17.

The school nurse in the same town would provide free condoms (and complete discretion) for over-fifteens. Clearly there is a strong case to be made that Lars' and Lisa's use of condoms benefits others? (See explanation in footnote.[1]) The benefits to other students (marginal social benefits) from the personal consumption (marginal private benefits) of condoms yield a societal benefit that is greater than the personal benefit to Lars and Lisa. There is a *positive externality*. Health care, education, libraries and civil defence are all examples of goods where there are benefits to non-users and/or non-payers.

Under- and overprovision of goods

As per the iteration on externalities above, goods which have negative externalities are over-provided. The market fails to provide the correct amount of the good since users do not see the negative effects on others. Tobacco, alcohol and petrol are all overprovided in terms of the actual costs to society in providing/consuming these goods.

When a good is under-provided, the market fails again. In a *perfectly* competitive market, young people will not buy health insurance, unemployment insurance or pension scheme. Nor would parents pay what it would cost to send their children to school or have street lights installed so they would be safer playing on the sidewalks. These goods have high external benefits to non-users but since the private consumer does not see these benefits he/she would under-consumer them.

Imperfect competition

When we speak of optimal resource allocation in competitive markets, we are assuming that there are a large number of buyers and sellers and that consumers have perfect knowledge and information about the price and quality of goods. In other words, a competitive market outcome is one where resources are allocated correctly in accordance with the price mechanism. Leaving this Disney-esque world of Mermaidomics, we find that firms are increasingly oligopolistic and that many of the goods we buy are indeed produced by a few large – often multinational – firms. You don't need a degree in rocket surgery to realise that if a firm lacks competition and can sell you 'less for more' then they will. This decreases consumer surplus (Chapter 8) and therefore leads to suboptimal resource allocation and market failure.

Common access resources

> "What is common to the greatest number has the least care bestowed upon it." Aristotle

I have a Rolex Explorer from 1959 in a safe in Sweden. In another box – and country – I have an Audemars Piguet and a Girard Perregaux. These are hideously expensive Swiss watches and they tick away like a dream whenever I have the fortune to take them out of their boxes. I have taken care of them for some 20 years now. When I stick my spoon in the wall (= die), Bell, Amanda and Johanna will inherit some wonderful timepieces for their children and grandchildren. Provided, of course, that they take care of these mechanical wonders as I have!

The hill overlooking Expat Street where I lived in Mexico is allegedly a lookout tower from the time of Montezuma some 500 years ago. Crumbling cement walls. Bags of trash strewn around. Old wire fences rusting away. Holes dug by locals needing land fill. Trees chopped down for firewood.[2] And of course the yearly forest fire which some idiot sets just for fun. Why am I not out on the hill picking up garbage and removing old fences? Well, it's not my land. Nor is it anyone else's. The result is the inverse of the Golden Rule; do it to others before they do it to you. If I don't chop some wood someone else will. If the hombres down the road don't dig the best gravel out somebody else will.

Resources that lack distinct ownership are said to be 'held in common' – nobody has the sole ownership and cannot deny others' access. Such common resources have been overused, destroyed and ultimately wasted. Examples of common

---

1 The answer is sexually transmitted diseases (STD) and teenage pregnancies. That's why the condoms were free. Many of my English colleagues are absolutely mortified when I tell them about this. I always ask them to look up teenage pregnancy and STD statistics for the UK and Sweden. None of them ever get back to me.

2 Rob and I *might* have had something to do with this. It was very cold. I point out that I am not *admitting* to anything.

resources include – but are not limited to! – rivers, oceans and air. All of these resources have historically and traditionally spanned across both private ownership and country boundaries whereby the *uncertainty or lack of property rights* has led to degradation and waste.

Priceless! Matt's trout stream. (Kölån, Helsingland, Sweden.)

Information asymmetry (HL)

My current car is an Audi A6 Quattro with a 2.7 litre engine and 250 horsepower. I bought it used and within three months I was reasonably certain I had been cheated. There was no way I was getting 250 HP out of this machine and the sluggish behaviour in traffic was a death trap – you can't be slow in entering a highway in Mexico when the cars are doing 120 kms per hour! Basically I bought a 'lemon' – an American term for crappy car.[3] Now, I spent some time checking the engine and even brought some knowledgeable colleagues with me to have a look at it before I signed the papers. Still got cheated. Why? Well, evidently the seller knew more about the state of the car than I did – he had far better information than I had. In this economic transaction the seller evidently had a greater power base in negotiating the terms of sale than I did – the information was not 'equally shared' but *asymmetrical*.

Commonly it is goods where sellers are knowledgeable due to the relative difficulties in acquiring adequate knowledge and information that are subject to this type of market failure. Used cars, shares and real estate are textbook examples of when the seller has an information advantage over the buyer in a transaction.

It should be noted that information asymmetry can work *against* sellers – for example when a somebody who has a serious disease buys life insurance. The insurer cannot in all cases guarantee that the buyer is telling the truth, so the insured person will in fact pay too little for the insurance. This is an example of *adverse selection*, where the seller simply lacks the information to correctly negotiate the 'correct' price of a good.

Summary & revision

1. Market failure: when the price mechanism fails to allocate resources optimally. The market fails when a good or service is sold in the wrong quantity at the wrong price. There are three main types of market failure:

   a. **Externalities** or negative spillover effects – when the MSC is greater than the MPC one speaks of *negative externalities* in production. When the MSB are greater than the MPB one speaks of *positive externalities*.

   b. **Under and over-provision of goods**. *Merit goods* such as education and health care and *public goods* such as bridges and streetlights are commonly under-provided whereas *de-merit goods* such as tobacco and alcohol are overprovided.

   c. **Imperfect competition** such as an oligopoly or monopoly market means the market has failed since more could be provided at a lower price without firms making a loss. (HL: The technical explanation for this type of market failure is that the price is higher than the marginal cost. See Chapter 27.)

2. Other types of market failure include

   a. **Common access resources** such as the ocean and forests. Lack of clear ownership rights means there is too weak an incentive for anyone to take care of these resources, resulting in over-use and thus sub-optimal allocation.

   b. **Information asymmetry** means that a party in an economic transaction has more/better information than the other. The result is that one party will get 'cheated' so to speak – a disadvantaged seller will set too low a price whereas a disadvantaged buyer will pay too high a price.

---

3 I urge you to look up George Akerlof (whose original name was Åkerlöf – i.e. of Swedish decent) and his 1970 work titled "The market for lemons".

17. Negative Externalities

Key concepts:

- Negative externalities
    - In production
    - In consumption
- Welfare loss
- Demerit goods
- Market based policies (and evaluation) to deal with negative externalities (taxation, tradable permits)
- Interventionist policies (and evaluation), government regulations, (age limits, licenses, govt control of sales, bans, licensing for firms)

We have expounded on how markets succeed in both the production and distribution of goods – and how the price function ultimately creates a market clearing. Recall that when quantity supplied equals quantity demanded, the market is in equilibrium and there is optimal resource allocation – a Pareto-optimum. Should – for some reason – the market not achieve this equilibrium then the market has failed.

Negative externalities

Externalities arise whenever the market clearing price creates benefits for or inflicts costs on a *third party* – i.e. a party other than the consumer or producer involved in the economic transaction. In the case of negative externalities, the *signalling function* fails to correctly portray the *real* costs to firms and consumers, resulting in over- production or over-consumption of goods. The market has failed to optimally allocate resources.

An easy way to get an immediate grip on the concept of externalities is to just think for a minute or two about the following question: Have you ever consumed a good which had a negative effect on others who didn't receive the good!? I warrant that your answer is 'yes'; how about cigars and noisy motorcycles as an example where other – non-consumers – bore the costs for goods they have not enjoyed? In this example, your *personal consumption* has had negative effects on others – external effects which we call a **negative externality** (or 'negative spill-over' effect).

Negative externalities in production

Conversely, when the marginal private costs (MPC) are outweighed by marginal social costs (MSC) then there is a **negative externality**. A good which renders higher marginal social costs than private costs is an example of overproduction; the production of coal-powered electricity causes emissions of sulphur and nitrogen dioxide which ultimately cause acid rain which kills forests and fish. Here, the market has underestimated the true costs of the good to society and produced too much of it.

> **Definition: 'Negative externalities in production'**
> A negative externality in production is a burden/cost endured by third parties arising from the production of the good. Negative externalities arise when the MSC > MPC.

Every year for over 25 years I have returned to my fishing paradise in Helsingland, middle Sweden, where my friends Östen, Musen and I have fished since pre-teenagerdom. Every year when I return, I have the same heart-in-throat experience of wondering whether the beautiful little brook-trout have finally succumbed to increasing pollution – primarily acid rain. You see, the middle part of Sweden is along a thermal weather path where a goodly proportion of emissions from Germany's industrial centre (Ruhrgebiet surrounding Köln) ultimately land. Sulphur and nitrogen oxides emitted in Germany ultimately land as weak solutions of sulphuric and nitric acid in Swedish lakes and streams. This lowers the pH-values of the

water and can rapidly kill off entire populations of trout and the insects upon which they feed.[1]

Therefore, when I drive the 850 km north in my German VW Polo, listening to CDs manufactured by the German firm TDK, and subsequently pitch my German VauDe tent by the river, I will **not** be paying the full price of these goods since the real cost – the social costs – are considerably higher! Even the people who have never had the thrill of stalking trout with a 9 foot fly-rod would seriously oppose measures to keep a natural strain of trout alive. Why? Because the cost would be… well, incalculable. How does one measure the loss of a sub-species?! Yes, one could measure the value of trout streams in purely monetary terms such as tourism and fishing license revenues, but there are numerous intangible benefits which, while impossible to calculate, are most evident; 'a living natural environment' sums it up. As Ronald Coase, one of the world's leading economists and Nobel Laureate in 1991 has put it; 'The question to be decided is: is the value of the fish lost greater than the value of the product which contamination of the stream makes possible?'[2]

The negative effects of pollution are shown in Figure 17.1. Using my fishing example, a tent-maker in Germany using Teflon[3] coatings does not include the costs of dying eco-systems in Sweden in the cost picture when focusing on maximising profits. Any level of output will have external effects on third parties – Swedish-American-Irish-Ecuadorian fly-fishing economics teachers on well deserved holidays for example. Output will be set at Q_0 (where **MPC = MSB**) rather than at the societal optimum of Q_{socopt} where **MSC = MSB**. This over-production of goods causes external costs to society – which will not be reflected either in the firms' costs or the prices paid by consumers.

1. **The extent of the negative externality:** The external costs to society – dead fish and very sad fly-fishermen – at Q_0 are shown as the vertical distance between the marginal private costs and the marginal social costs; **MPC ⇔ MSC**.

2. **Welfare loss:** The shaded triangle shows how the **MSC** is higher than the **MSB** for any unit of output produced between $Q_{soc\,opt}$ and Q_0 – this is the welfare loss (burden) to society in producing these units.

Welfare loss in producing $Q_{soc\,opt}$ to Q_0 more units of plastic – each unit has a higher MSC than MSB

MPC ⇔ MSC; extent of the negative externality in production

Figure 17.1 Negative externalities of plastic production

The MSB is assumed to be the same as the MPB, i.e. all benefits to society are accounted for by the private benefits. Price and output is set where private costs equal private benefits. Since third parties suffer from this, there are negative externalities arising from consumption. (Social costs = private costs + external 'spill-over' costs.)

Please note that in the examples used, acid rain, carcinogens and particulates, will naturally have numerous negative externalities other than for us fishermen. I remember when one could walk around hundreds of statues littering the grounds of the famous Acropolis in Athens. Returning some 30 years later, I found that most of them have been moved indoors. While this might in part be due to the risks of theft and vandalism, it is often simply to protect the invaluable antiquities from the highly damaging acid rain and air-born particulates resulting from heavy traffic and industries – which have eroded this

---

1   Local municipalities and fishing clubs combat this by dumping large amounts of limestone in source lakes and feeder-streams in order to raise the ph-value.

2   *Coase, Ronald; The Problem of Social Cost, Journal of Law and Economics,* October 1960, page 3.

3   Teflon contains perfluorinated chemicals (PCOs) which accumulate in protein and can build up to high levels in both humans and wildlife. Recent studies suggest PCOs are linked to birth defects, developmental problems, liver damage and affect the neuroendocrine system. (I am *way* out of my depth here and had to look this up…so you better ask your biology teacher.)

priceless cultural heritage more in the past 40 years than the preceding 2,000 years!

DANGER MINES WARNING !

Frequently students fail to differentiate sufficiently between 'extent of the externality' and the 'welfare triangles'. The externality shows the net benefit/cost to society in the consumption of the good. The triangles show how the external effects result in misallocation; a potential welfare gain or welfare loss. I also urge you to 'pull out the arrows and triangles' as I have done in the illustrations in order to make it easier on your examiner.

The concept of externalities can be applied quite broadly. A new building which blocks your view confers a loss of utility on you – and thus you, as a third party, will bear the cost of lost viewing pleasure. This is a negative externality for society. A new building which enhances your viewing pleasure will of course bring society positive externalities. It goes without saying that one ultimately stands on a pedestal of normative judgement in assessing external costs and benefits; a free pop concert in the park will bring both depending on one's preferences. It is obvious, however, that there are some costs which are more objectively observable. An example is a World Bank study in 2000 which attempted to assess the social costs of air pollution in developing countries.[4] The study showed that air pollution could be linked to the premature death of over 500,000 people each year, costing some 2% of GDP in health care and lost output.

Negative externalities in consumption

Here's the one you've all been waiting for; drugs. Nothing quite like quaffing a few pints whilst puffing on a sizable Havana cigar.[5] Oh, you were perhaps thinking of *illegal* types of mind-altering substances, say cocaine and LSD. We'll get to those, but for the time being I shall deal with the more conventional (and in western countries, socially acceptable) substances alcohol and tobacco.

Author and victim – he'll no doubt sue me in 20 years if I'm still alive.[6]

Demerit goods

Now, when I sit out on the balcony smoking my cigar and having my evening Cognac, my lady Bell immediately closes the balcony door – forcing me press my nose against the window in order to catch the news on TV.[7] She still loves me…but not the smoke I exhale! She, the self-professed economics ignoramus, is highly aware of the damage to others caused by second hand smoke – and just as aware of the negative effects my consumption will have on me in the future. She rather wants me around for a while longer.

It is often said by my more socially aware colleagues, that if cigarettes and alcohol were previously unknown and then invented tomorrow, they would immediately be banned and severe sanctions for selling and consumption would be imposed! They are quite right, to the extent that a lawsuit in Florida against American tobacco companies in 2000 resulted

6 Before I start getting hate mail, the story goes something like this: Summer of 2008 I was back in Sweden for a quick visit to Guy and his family – basically my adopted family. His (my) sister Monica had adopted two wonderful Korean lads and standing on the lawn of Musen and Pernilla's house it was time for a picture with Uncle Matt. I'm not too comfortable with children and the discussion prior to the picture went something like this: "Matt! Picture time – get in there with the kids!" "Em, I'm smoking a cigar." "So put it down!" "I'm busy! It's a good cigar…" Someone tossed me an infant…and before I knew it, Toni the bendejo snapped a picture.

7 Yes. I'm getting a portable TV for the balcony cigar and Cognac sessions. Revise 'complement goods' right about here!

---

4 http://www.worldbank.org/cleanair/global/topics/health_imp.htm

5 These are, of course, **complement goods**.

in a ruling that \$US145 billion (145,000,000,000) be awarded to smokers (in Florida alone!) for health damage.[8] While this might seem a trifle on the excessive side, no-one can possibly deny over 50 years of research showing very high correlation between smoking and various types of cancer, lung disease and premature death. Alcohol might well be even more costly for society; not only do we have to contend with the severe health damage to individuals but also with the loss of labour skills and output resulting from inebriation, hangover, alcohol related disease, and, again, premature death.

It's simple: don't drink and drive.

Both tobacco and alcohol are **demerit goods**, since users either ignore or do not realise the full costs of consumption – both to themselves and third parties. Both goods show the characteristics of the individual consumer not being able to see how his/her consumption patterns negatively affects society. Indeed, when it comes to alcohol and illegal drugs, a case can be made for judging many individual consumers as seemingly incapable of making a correct decision even in terms of their own self interest!

> **Definition: 'Demerit good'**
>
> A good which has negative externalities and where users do not realise – or ignore – the costs to themselves and the societal costs inflicted on third parties. Common examples are tobacco and alcohol.

The negative effects of consuming tobacco are shown in Figure 17.2. My consumption of tobacco will create costs for non-users and me in the future – my health issues will of course inflict costs upon society such as hospital care and the loss of labour hours. Output will be set at Q_0 (where $MSC = MPB$) rather than at the societal optimum of $Q_{soc opt}$ where $MSC = MSB$. This over-consumption of a demerit good (tobacco) causes *external costs* to society – which are not reflected either in the firms' costs or the prices paid by consumers.

1. **The extent of the negative externality:** The external costs to society – health issues and loss of income due to increased sick days – at Q_0 are shown as the vertical distance between the marginal private benefits and the marginal social benefits; $MPB \, \acute{o} \, MSB$.

2. **Welfare loss:** The shaded triangle shows how the MSC is higher than the MSB for any unit of tobacco consumed between $Q_{soc opt}$ and Q_0 – this is the welfare loss (burden) to society in consuming these units.

Welfare loss in consuming $Q_{soc opt}$ to Q_0 more units of tobacco – each unit has a higher MSC than MSB.

$MSB \Leftrightarrow MPB$ extent of the *negative* externality in consumption
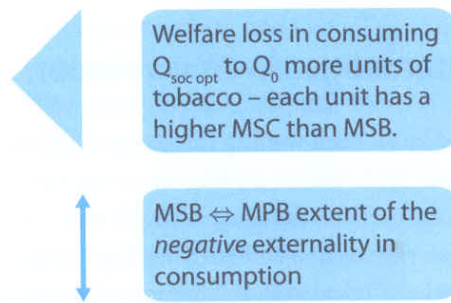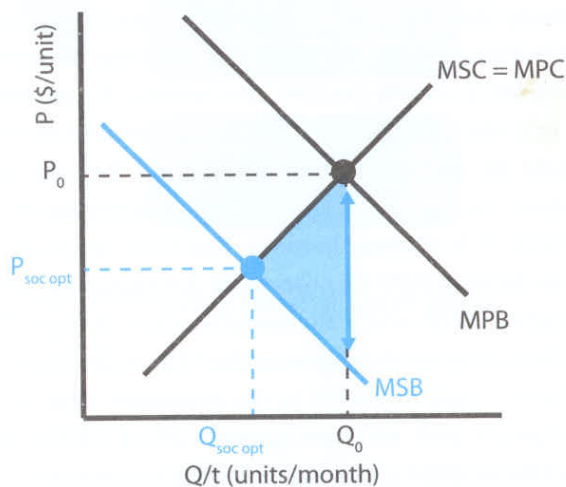
Figure 17.2 Negative externalities of tobacco consumption

The **MSC** is assumed to be the same as the **MPC**, i.e. all costs in production are accounted for by the firms. Price and output is set where private costs equal private benefits. At P_0 there is an overconsumption of $Q_{soc\ opt}$ to Q_0 since every one of these units of tobacco has a MSC greater than the MSB. This is the shaded triangle representing the welfare loss to society.

Policies to deal with negative externalities in production

Figure 17.3 exemplifies a good which creates a negative externality in production; MSC > MPC and output is Q_0 rather than the allocatively efficient level of $Q_{soc\ opt}$. (Figure grossly exaggerated for clarity.) In order to correct this market failure, the government imposes a (unit) tax of, shown by the double-sided black arrow between the MPC_0 curve and the MSC curve. (Note how we assume that the 'money' cost of smoking or such can be estimated correctly. In fact, this would be virtually impossible in reality, as assigning a money cost to loss of quality of life would be both highly subjective and arbitrary.) The tax serves to shift the MPC_0 curve left to $MPC + tax$, which lowers the quantity demanded to Q_1. The results are a) lower externalities, and b) lower welfare loss to society.
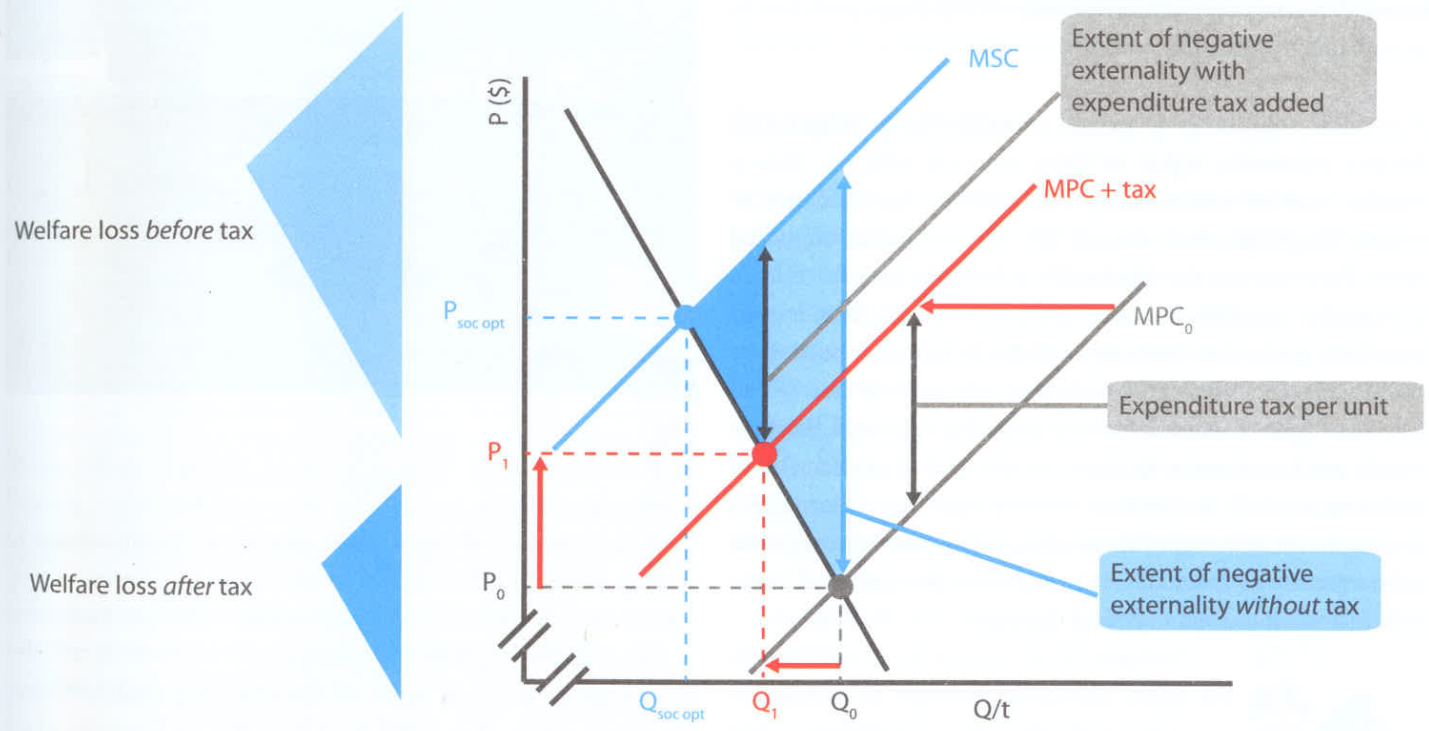
Figure 17.3 Taxing away negative externalities of production

In taxing the production/consumption of the good, the real costs have been brought forward and now burden those who are responsible for the externalities. Note that taxes can be imposed on either the consumer or the producer, such as a tax per pack of cigarettes or a pollution tax on emissions by firms. Both forms serve to *internalise* the externality. Smokers will bear more of the real cost of their habit while in pollution cases it is an example of the 'polluters pay principle', which acts as a disincentive to producers to continue to pollute. Additionally, in a best-cast scenario, the government might divert tax revenue to those directly affected by the initial externalities, such as using road taxes to build sound barriers and plant trees between major roads and housing areas. This would help to correct the inequity of third parties paying a cost which rightly belongs to users/providers.

Two main problems in taxing away externalities arise: the first is the difficulty in deciding *how large a tax* to set on output. It is almost impossible in many cases to correctly estimate the extent of the negative externality (e.g. "...what is the real value of an indigenous trout species in Helsingland, Sweden") and thus impossible to estimate the 'correct' level of taxation. The next problem arises in *attributing* the pollution to firms – and a third issue arises when offending firms are on the other side of a national border where the domestic government has no control!

Internalising externalities

One of the most common ways to bring external costs back into the fold of the market place is to levy costs on firms producing goods which have negative externalities – the *polluter pays principle*. Commonly, such goods are associated with external costs in production, such as circuit boards where ozone-depleting chlorofluorocarbons (CFCs) are used in the manufacturing process.

A possible solution is to extend property rights by granting distinct ownership rights to those who are affected. This is another way of **internalising** externalities; any damages to private property create a basis for claims on the offending party. By doing so, the externality is brought into the fold of the market mechanism since previously un-attributed costs have been made clear and responsibility is assigned. In the case above, citizens suffering from prematurely rusting cars would have the right to claim from the polluting firm and the firm would pay for property damages. In the case of air/river/noise pollution we will get market failures and thus externalities since nobody owns any of these objects. One way to correct the failure would be to establish firm private ownership of them and thus allow owners to collect damages.

Yet again, collecting damages in a situation where the offending firm is far away and cannot be directly connected to the car owners' loss of property is almost impossible. I didn't choose this example arbitrarily, as I simply wish to show that in many situations it is impossible to allocate and much less enforce property rights. However, let us look at two simplified examples of how an extension of property rights might be successfully implemented.

The Beach – how to exert pressure on offending firms

Imagine a beautiful beach in the summer vacation paradise of your choice. Sand, cloudless 32 degrees and nice breeze, crystalline azure blue water...no, wait, isn't that raw sewage pouring out into the bay?! Yech! And the sand resembles used kitty-litter upon closer examination. No wonder the tourists are at the beach front restaurants and cafés instead. Or rather, 'aren't' . It seems that the pollution has driven off a good many tourists and the restaurants and cafés are not as full as they once were. This is a clear case where other parties – manufacturing industries, households, hotels etc – have inflicted damage on others who have a clear interest in clean beaches; the beach front businesses.

A possible solution here is to grant property rights to all businesses along the beach front. Give each business a granted right to a section of beach front and allow the businesses to extend into sun-chair rentals, windsurfing schools, diving classes and the like. This creates an incentive for the businesses to clean up the beach as they compete for tourists. It will also induce them to put pressure on the offending polluters since the externality is damaging their property and thus in effect harming their profits. One could expect the beach businesses to act towards either implementing new legislation on sewage or demanding that existing suchlike is enforced. It is conceivable that the community acts to charge households and industries for increased water purification and/or increase taxes in order to build cleaning plants, since the costs of the externalities such as loss of tourism profits and tax revenues will be made apparent by the new property owners.

MICROECONOMICS

The River Runs Through It – how to get offending firms to exert pressure on themselves

What if a factory upstream from where I fish trout were to release effluent into the river causing the trout population to decline over time? This would be quite a blow to the thousands of fishermen who regularly return to the stream. They would bear the external costs of the factory's activity.

A possibility is to extend the property rights of the firm to encompass the entire river! Say that the municipality were to grant the firm sole entrepreneurial rights along the river. River rafting, fishing, camping – you name it. The decision the firm would face is whether the river is more valuable clean or dirty; in other words, would the marginal revenue of all the river sports activities be greater than the marginal costs involved in reducing pollution to acceptable levels? This solution would of course only be considered if the answer were 'yes', and therefore the firm would be able to add to total earnings by keeping the river clean. In creating a framework where the party responsible for a negative externality (cost) must take them into account, the externality has been brought into a market framework, i.e. internalised. However, it is quite possible that the original property right owners – say fishing clubs – might simply have to pay the firm to limit pollution.

Definition: 'Extension of property rights'

Extending property rights means there is an explicit right of use of an asset or resource. Ownership – and thus responsibility – is granted, creating an incentive for owners to take into consideration all costs, i.e. externalities. In this way private costs are brought closer to social costs and firms are more inclined to limit total costs in line with a profit incentive. The externality is in this way internalised by bringing external costs into a market-based framework.

It should be noted that the concept of expanding property rights is quite wide. Many examples involve granting sufferers the right to litigate (= take legal action, sue) against offenders. For example, many cases of lung cancer have been brought within the legal jurisdiction (= responsibility) of firms which produced asbestos in the 1960s – firms will have to compensate these workers as the workers' ownership of factors of production (= their own labour) has been harmed. The workers' damaged property will have to be paid for by the offender – the firm. A most notable case of late, is the class action suit (where many people who have been injured get together and sue a company as one) against the Swiss/Swedish multinational engineering company, ASEA Brown Boveri (ABB).[9] When ABB bought the American company Combustion Engineering in 1989, it also took over all of the American company's liabilities – even in terms of previous employees who claim damages from dealing with asbestos. ABB is expected to pay over USD1.2 billion in settlements for claims and total costs together with legal fees could run into the billions.[10]

While assigning property rights encounters any number of problems, such as **who should pay** whom and how much, there is a market-based appeal to the system. There is no need for extensive regulatory bodies which monitor and estimate the extent and cost of the externalities. This is instead done by the relevant parties; the offending firms and the third party sufferers. In setting out well-defined property rights (which is obviously the tricky part!) the offender and victim will be able to arrive at equitable compensation which optimises resource allocation and minimises environmental damage.

Tradable permits

While this might sound like the medieval custom of sinners buying indulgences, a permit to pollute is slightly different – but almost as controversial. This highly market-based solution to pollution is the establishment of programs wherein a government regulatory agency sets a ceiling on total emissions and then emits tradable '**pollution permits**' to offending firms. A tradable permit scheme essentially means that government decides the 'acceptable' level of pollution and then issues credits or allowances to firms and fines firms which pollute more than their permits allow. The permits are allowed to be re-sold to other firms, which in theory – and in practice in some cases – creates both an incentive to clean up production and a disincentive to exceed the limits imposed.

How tradable permits work

Consider a market comprised of ten firms, producing a market total of 100,000 Widgets and also emitting a total of 1,000 tonnes of *greenhouse gases* (**GHG**, mainly carbon dioxide, ozone, methane, chlorofluorocarbons and nitrous oxide). Let us for the sake of simplicity say that this comprises the sum of externalities, which means that each Widget has an external

---

9 An attempt to sue McDonald's for causing obesity was tossed out of court in the US in 2003.

10 http://www.forbes.com/2003/01/17/0117autonewsscan01.html and http://www.corpwatch.org/news/PND.jsp?articleid=1888

MICROECONOMICS

cost to society of 10 kg of GHG per Widget (1,000 tonnes of GHG / 100,000 Widgets). The government sets a ceiling of 500 tonnes of GHG and initially allocates 50 one-tonne pollution permits to each firm. Assuming that this would also halve market output, 50,000 Widgets would be produced. What happens next depends largely on how efficient each individual firm is in producing Widgets, but here is a possible scenario:

1. Say that the three *most polluting firms* (= most inefficient firms) can only produce 1,000 Widgets each when limited to 50 tonnes of GHG emissions and would have to invest so much in new filtering technology in order to produce more that it simply wouldn't be worth it. These high-polluting firms leave the market but hold permits allowing for the emission of 150 tonnes of GHG which they will not use. Total market output of Widgets is now 47,000.

2. The three exiting firms will sell the permits on the open market. The question being what will set the market price? Simple; the foregone profits of firms not using the permits will set the floor while the highest bidders, which will be the *most efficient firms*, will set the ceiling – since these firms will be able to produce the most Widgets per tonne of GHG and thus be willing to pay the most for the permits. (Keep in mind that the market price for Widgets will rise as three firms have left the market.) Say that five of the remaining firms are producing half of their previous output and are marginally efficient, i.e. they have no incentive to buy additional permits. They produce 25,000 of the market total of 47,000.

3. Consider now that the final two firms, YellowFirm and GreenFirm, are also the most efficient firms and are bidding for the extra 150 tonne allotment. YellowFirm is producing 10,000 Widgets with its 50 permits and estimates that it can produce an additional 20,000 Widgets with the additional 150 permits up for sale. (Note that the additional 150 permits does not mean that the firm can increase output by 30,000 units! This is of course due to increasing marginal costs.) At the going market price of €10 per Widget, YellowFirm's additional revenue of €200,000 (€10 times 20,000 more units) would cover all additional costs – both production costs and permit costs – if the price of the permits were maximally €500 each.

4. GreenFirm is producing 12,000 Widgets and could produce an additional 30,000 Widgets with the 150 permits. The marginal revenue of €300,000 would cover

all additional costs even if the cost of the permits were as high as €510 each.

5. GreenFirm wins the bidding war, and after taking on Bell (153 centimetre ferocious Australian female) as chief negotiator, pays €501 per permit.[11] The final outcome is that the five non-bidding firms produce 25,000 Widgets; YellowFirm produces 10,000; and GreenFirm 42,000. The **sum output total of 77,000 Widgets** results in **500 tonnes of GHG**, or unit external costs of **6.5 kg of GHG per Widget**.

6. The short run result is that pollutants contributing to the green house effect have decreased by 50% while output has decreased by proportionately less, namely 23%.

7. It might appear that a polluting firm would have no real incentive to lower pollution levels once the permits have been purchased. This is faulty economic reasoning, as the firm has resources (money) tied up in the permits and thus incurs an opportunity cost. In the long run there will be an incentive for firms to increase efficiency and lower emissions in order to free resources by selling the permits. Put more bluntly; any new production method or technology which can save a polluting firm one tonne of pollution at a lower cost than buying a permit on the pollution market will be attractive. It will also enable firms to sell off unneeded permits and free up funds for other investment. (See below; *Case study; CO_2 permits and the Clean Air Act (USA)*).

The system does have a number of **disadvantages**, however. Many environmental groups feel that the market based system sends the wrong signal to both producers and consumers, in that there are 'permissible' or 'non-harmful' levels of pollution. The system might simply mislead society into believing that pollution is decreasing when in fact it is not – only the pollution per *unit*. It is also feasible that heavy polluters buying up ever-cheaper permits due to the diligence of firms cleaning up their act might actually perpetuate the very pollution the permits attempt to reduce. A rather tricky issue is also that pollution knows no boundaries – if the USA has strict controls via permits and Mexico does not then firms might avoid the issue by establishing subsidiaries right across the border in Mexico. And finally, quite frankly, most economists seem to favour the appealing simplicity and

11 Woe betides the hotelier or restaurateur who tries to do a number on her!

straightforwardness of emission taxes rather than what often turns out to be rather complex systems of pollution permits.

Policies to deal with negative externalities in consumption

However high the costs to society may be due to the production of goods, many negative externalities arise from the *consumption* of goods such as alcohol, petrol and cigarettes. In such cases, taxing the good via *expenditure* is the basis of an incentive-based solution to distortions arising due to market failure. The cardinal rule of taxes on expenditure (see Chapter 13) is; 'When you want less of a good, put a tax on it'. The other rule deals with consumers' preferences and marginal utility; if you want people to consume less, adjust their demand by way of changing the way they think.

Figure 17.4 shows the effects of government intervention in the market for a good which has negative externalities in consumption; alcohol. Market equilibrium at P_0 and Q_0 render private benefits that are clearly beyond the societal benefits ($Q_{soc\,opt}$), resulting in a negative externality of MSB ⇔ MPB. This is not rocket science; street violence as drunken lads spew out of pubs at 01:00 in the morning…traffic fatalities due to drunken drivers…loss of income and therefore tax revenue as workers stay home to nurse hangovers…and the loss of human capital due to premature death resulting from the consumption of alcohol. The over-consumption of $Q_{soc\,opt}$ ⇔ Q_0 is illustrated by the light blue triangle – every one of these units has a higher social cost than social benefit and is thus a welfare loss.

By intervening on the market, a government can reduce the supply of alcohol (MSC_0 to MSC_1 in Figure 17.4) or decrease demand (MPB_0 to MPB_1 in Figure 17.4). In both cases the quantity consumed and market price move closer to the societally optimum levels and the welfare loss is decreased. Governments use a number of methods to limit the supply of alcohol (apart from the most obvious which is a ban):

- **Decreasing supply:** A *tax on alcohol* increases the costs for firms and decreases supply as do government limits to the amount of licensed alcohol outlets allowed. (In many notable cases such as Iceland, Sweden, Norway and Finland, all sales of alcohol are sold solely through a government monopoly.) *Limited opening hours*, for example not selling alcohol after certain hours and weekends (Sweden).

- **Decreasing demand:** For alcohol – and the highly non-complementary good driving – the issue is to get consumption down by *advertising negative aspects of consumption*. Anti-drinking campaigns hope to decrease demand (MSB_0 to MSB_1 in the top diagram) and move consumption levels towards the social optimum at $Q_{soc\,opt}$.[12] Age limits for under-18s (UK) or under-20s (Norway) also limit demand for alcohol as do warning labels and strict rules on the advertising of alcohol.

MSB ⇔ MPB extent of the negative externality in consumption

Welfare loss in consuming $Q_{soc\,opt}$ to Q_0 more units of alcohol – each unit has a higher MSC than MSB

MSB ⇔ MPB₁ extent of the negative externality in consumption after government policies reducing demand/supply

Welfare loss in consuming $Q_{soc\,opt}$ to Q_0 more units of alcohol after government policies reducing demand/supply

Figure 17.4 Dealing with negative externalities of alcohol

---

12 We were driving on the highway outside of Windhoek, Namibia when a most eye-catching sign appeared by the side of the road. Apart from the less-than-subtle sign itself, which read 'Drink kills', there was an additional extra to really drive the point home, so to speak. A smashed-up car had been wrapped around one of the poles holding the billboard up. A most effective message. Another ad we saw frequently in Namibia was an AIDS poster where a young lady says 'My boyfriend said that using a condom takes all the fun out of it. Boy did I have fun proving him wrong!'

The objective is to alter peoples' behaviour. Making a good harder and/or more expensive to obtain decreases consumption by decreasing supply. A major problem here is that the price elasticity of demand for demerit goods is low – primarily due to the lack of substitutes. There is also the black market effect; high taxes make it profitable for illegal sales of alcohol and tobacco – some studies put forward that the resulting criminality actually imposes even greater costs on society than legal consumption of the good would! (See for example the on-going debate in the US and Mexico as whether to legalise drugs.)

Another issue is how to tax a harmful good without having numerous negative spin-off effects. One third of US citizens are now clinically obese costing some USD147 billion in obesity related medical care and the US has debated for years on the possibility of a 'fat tax' aimed at reducing the consumption of sugary and fatty foods – 'junk food' in daily slang.[13] The difficulty is that in aiming for, say, a decrease in the consumption of sugar via a tax, firms producing ethanol or using sugar as a flavour enhancer in food production will see costs rise. Also, which goods are healthy and which are unhealthy – a banana or grapes might well wind up in the same tax bracket as a chocolate bar! A key issue with economists is also that a tax on sugary/fatty foods becomes highly regressive – lower income groups will pay a larger proportion of their income in tax.

As for government policies aimed at demand, educating and informing citizens of the real costs of use can change market demand and thereby correct over-consumption of goods. Yet this is often a long term plan over generations as evidence suggests that anti-drinking campaigns and 'health awareness blitzes' have limited longevity in the hearts and minds of consumers. Societal mores and traditions take a long time to change it would seem.

13 Wall Street Journal; *Cost of Treating Obesity Soars*, July 28 2009. To date only Norway, Finland and Hungary have introduced some form of 'fat taxes' on pre-packaged foods with high salt or sugar content.

Summary & revision

1. **Negative externalities in production** occur when firms do not see the costs to third parties and thus over-produce the good in terms of societal welfare. Common examples of negative externalities in production are air pollution, water pollution, noise pollution and deforestation resulting in soil erosion and/or desertification.

   a. The **extent of the externality** is the difference between the marginal social costs and the marginal private costs (MPC > MSC)

   b. The **welfare loss** is the allocative loss in overproducing a good. The sum of all units produced where the marginal social cost is greater than the marginal social benefit (MSC > MSB) shows the welfare loss to society.

2. **Possible solutions** to negative externalities in production are:

   a. Expenditure **taxes** on the good

   b. **Polluter pays** principle – property rights are assigned and offenders must pay for the damage caused

   c. **Extension/allocation of property rights** giving firms a vested interest in taking care of the property

   d. **Tradable permits** – maximum levels of pollution are established and 'pollution permits' can then be bought and sold on an open market. The aim is that the most efficient firms (= least polluting) have an incentive to pay the most for additional permits.

3. **Negative externalities in consumption** occur when consumers do not see the costs to third parties and thus over-consume the good in terms of societal welfare. Common examples of negative externalities in consumption are the health and social effects of tobacco and alcohol (e.g. demerit goods).

   a. The extent of the externality is the difference between the marginal social benefit and the marginal private benefit (MSB < MPB)

b. The welfare loss is the allocative loss in over-consuming a good. The sum of all units produced where the marginal social cost is greater than the marginal social benefit (MSC > MSB) shows the welfare loss to society.

4. **Possible solutions** to negative externalities in consumption are:

a. Decreasing supply by way of expenditure **taxes** on the good

b. Decreasing supply via **regulation** of opening hours, access to the good by way of age limits

c. Decreasing demand by using **negative advertising** and awareness campaigns

18. Positive Externalities and Merit Goods

Key concepts:

- Positive externalities
  - In production
  - In consumption
- Merit goods
- Government policies for merit goods

MICROECONOMICS

Positive externalities

As soon as there are third party benefits in production/consumption, we speak of positive externalities. Society benefits when I hire workmen to beautify my house and garden since this leads to an increase in the value of my property and can therefore increase property values in the neighbourhood. Society benefits also when I take time out from writing to go for a run – my increased physical fitness means I can work harder and longer and will have an improved immune system and thus fewer sick days per year.[1]

Positive externalities in production

Let's get one of my favourite complaints out of the way right now; the issue of positive externalities in production is hideously contrived and most of the teachers I've questioned about it resort to "…honey manufacturers keep bees which then pollinate neighbouring apple trees…". Great. Anything else? Well, not much as it turns out. In fact, most of the examples I have come across could equally as well wind up under the next heading 'Positive externalities in *consumption*'. The problem lies in that for third parties to benefit from the *production* of a good, the positive spill-over effects must exist whether the good is subsequently consumed or not! Few goods measure up to this description as it turns out.

One example I believe *might* measure up to the demand of '…my production benefiting you – even if nobody consumes my output…' was given to me by my remarkable editor, Rory, during one of my stays with him in Melbourne. He told the story of a landowner out in the country who had been brought up on charges of deforestation of state land. The landowner's defence was that he had simply been producing fire breaks (corridors kept clean of trees and brush) in order to protect his house from the frequent forest fires that raged in the area. When a subsequent fire during his trial burned down several homes in the region but spared numerous houses adjacent to the landowners' fire breaks he was exonerated (= cleared) on all charges.

Regrown Australian bushland ... "...down by a billabong..."

---

1 Though having just tried out my new 10 kg weight vest on a 6 km run I'm not so sure anymore. I think I went past the marginal net *benefit* stage and hit net *negative* effects…the thought occurred as I was collapsing from heat exhaustion in a nearby banana plantation here in Indonesia.

MPC > MSC; extent of the positive externality in production

Potential welfare gain in producing Q_0 to $Q_{soc\,opt}$ more kms of firebreaks – each unit has a higher MSB than MSC.
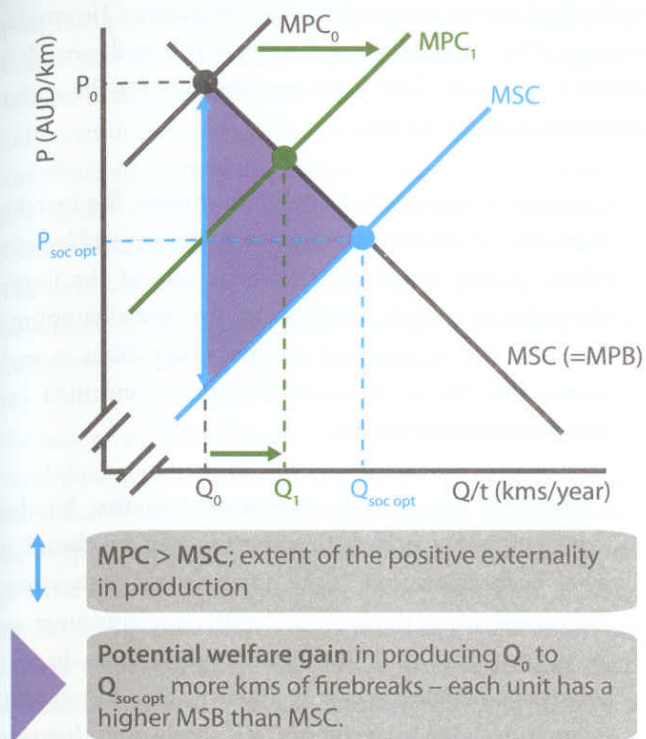
Figure 18.1 Positive externalities of fire wall production

Figure 18.1 above shows how the costs to a single landowner (MPC_0) are higher than the costs to society (**MSC**) in producing fire breaks. This is the extent of the positive externality conferred upon other landowners enjoying increased protection from forest fires without having paid for it. The welfare triangle shows all the kilometres (Q_0 to $Q_{soc\,opt}$) of additional fire breaks that would have higher social benefits (**MSB**) than costs (MSC_0). The curve MPC1 shows the effect of government attempts to approach the societally optimal level ($Q_{soc\,opt}$) by way of direct provision or subsidies. See below.

Another reasonable example of positive spill-over effects in production is wind and solar powered electricity. By putting up windmills to capture wind power and solar powers to harness solar radiation, a certain 'saving' might occur in the building of new coal fire plants. In other words, there is a *savings* of negative externalities in the production of clean alternatives. Having said this; it is obvious that there are no *direct* positive externalities in putting up windmills!

Positive externalities in consumption

A *merit good* is a good where there are positive externalities to such an extent that society deems the good to be *under-provided* by free market forces. (See further on.) Subsequently

the good is also under-consumed, since consumers lack sufficient information to consume the 'correct' amount. As shall be explained, merit goods such as education, health care and pensions are considered highly beneficial to society and are therefore in many countries supplied by public monies. It is viewed as socially beneficial for more of the good to be produced and consumed than would be supplied by a competitive market.

Let's use education as an example. Most consumers (students) would hardly be inclined to pay the full cost of education since they would not take into account future personal benefits. Nor would the person enjoying (?!) economics lessons be the only beneficiary of consumption since society at large will ultimately gain from a better educated labour force. The presence of large positive externalities moves many countries to provide the goods directly by running state schools.[2] (In fact, few – if any – schools are able to survive on *purely* private tuition, and most private schools will be subsidised in some way by grants and contributions from private firms, ex-students or governments.)

Figure 18.2 shows the positive spill-over effects in consuming education. At market equilibrium (P_0 and Q_0) the positive externalities (to individual users in the future and society in general) would be such benefits as increased lifetime incomes, greater enjoyment of art and literature and the increase in tax revenues paid by educated people earning higher wages. The societal benefits are shown by the MSB curve and the double-edged blue arrow shows the extent of the positive externality. Societal optimal point of output is where **MSB = MSC**; Qsoc opt. Since society is under-consuming and under-providing education by the amount of $Q_0 \Leftrightarrow Q_{soc\,opt}$, each of these non-provided units has a higher social benefit than social cost, e.g. the **MSB > MSC**. This is the potential welfare gain in producing $Q_0 \Leftrightarrow Q_{soc\,opt}$ more hours of education per year.

The two green curves – increase in supply (MSC_1) and increase in demand (MPB_1) – show the possible effects of government policies and are dealt with further on under "Government policies…".

---

2   Again, be wary of confusing 'provision' with 'supplied by'. Many countries increasingly rely on various forms of 'school vouchers' where consumers (parents/students) can spend their allotment on the school of their choice. Thus, it is possible for a student to be educated by a private facility using public money.

Extent of positive externality

Potential welfare gain
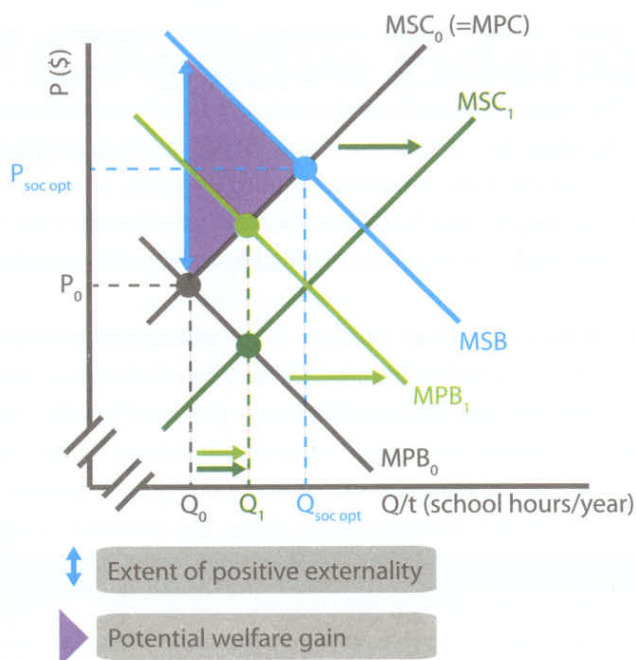
- Government **subsidies** or government provision increase supply from **MSC** to **MSC+subsidy**, bringing total output to Q_1 and closer to the societally optimal output of $Q_{soc\,opt}$.

- Government **incentives** such as paying families for every day their children are sent to school will increase demand for education, shown by the shift from MPB_0 to MPB_1 – again, output is closer to the societally optimal level of $Q_{soc\,opt}$.

Figure 18.2 Positive externalities of education

Merit goods

The basic issue of a **merit good** is that it is *meritorious* (Did I tell you to 'taste the term' yet? Oh, right.) to society as a whole. One can define a merit good as one where the consumption of the good is highly beneficial to society yet would be gravely under-consumed by the individual – i.e. the MSB > MPB. Some classic examples are **education, health care** and **pension schemes**. Merit goods, in spite of having benefits for both users (in a foreseeable future) and non-users, will be under-consumed in a competitive market since users cannot see how their use benefits others – or themselves in the long run.

So, why would we not be inclined to consume 'enough' (as defined by the societally optimal level of MSB = MSC in Figure 18.2) of these goods – especially when economics shows us how society as a whole receives positive external effects? Why would people *under-consume* IB physics or French lessons and health care? Simple, the short-sightedness and self-interest of

the individual cannot foresee how consumption of Newtonian physics and TBC vaccinations will ultimately benefit oneself in the future – or others. Two main reasons exist for the under-consumption of merit goods:

1. **Consumers ignore the benefits to others:** The inability to see how personal consumption spills over and benefits other societal members is perhaps one of the largest obstacles in people consuming the socially optimal quantity. The existence of *positive externalities* in itself means that the good would be under-consumed and therefore underprovided.

2. **Consumers ignore their own future benefits:** Another explanation for under-consumption can be found in what economists call '**time inconsistent behaviour**'. This arises when the benefits to the user/consumer are far in the future while the costs are immediate. In such situations, people are not inclined to give the benefits enough weight in weighing the costs and benefits. Giving up present enjoyment in order to learn French would increase possible future benefits in travel and employment possibilities – but how many would actually think along these lines?![3]

> **Definition: 'Merit good'**
> A merit good is a good where individual consumption benefits the user in the future and others in society, e.g. there are positive externalities. It is commonly under-consumed/under-provided due to consumers underestimating or ignoring the long run personal and societal benefits. Examples include health care and education.

Health care and pensions

Let's use two classic examples of merit goods; **health care** and **pensions** – both have strong elements of positive externalities and time inconsistent behaviour. Most European countries have relatively advanced publicly funded health care and pension systems – rather than a system based on households paying to a private health insurance or pension fund. The reason is that

---

3   One of my favourite examples of time inconsistent behaviour was pointed out by one of my students. In the movie '*Bridget Jones's Diary*' there's the scene where she writes; 'Monday 28th of April. Gym visits, zero. Number of gym visits so far this year; 1. Cost of gym membership per year; £370.'

most European governments have a strong element of social-welfare woven into the ideological policy base – regardless of political hue. Europeans have long since considered it wasteful and harmful for the health-care system to be based on private contributions. Simply put, people in their 20s to 50s would not give health problems (resulting from aging and hard work, perhaps) and retirement benefits enough thought and would thus have a strong propensity to neglect putting aside income for *possible* ill-health and *future* old-age. ('Live fast, die young and leave a beautiful corpse', and all that.[4])

The lack of adequate health-care will ultimately lead to illness – and thus to sick-leave and loss of possible output. The private individual loses good health and income. Society loses both the output, income and tax revenue which could have been generated through health-care. As for pensions, it has been proven that people on lower incomes have a far more marked tendency to neglect pension fund payments, and thus will ultimately suffer from considerable declines in income when they retire. The *under-consumption* of both these merit goods will lower both life expectancy and the quality of life. Both represent a societal cost which European governments have long considered too high to bear and therefore use tax revenue to fund them and many other social/welfare goods.

For all you young people out there…

One final example, just for you teenagers! In class, I often bring up condom use as a very clear example of where individual use confers very high external social benefits – not only does the private individual benefit from protection but society at large benefits from the decrease in the spread of sexually transmitted diseases (STDs). The aim, obviously, is to get people to consume more of the good in order that society benefits from lower spread of venereal diseases – not to mention the societal costs of unwanted children, and in many LDCs, population growth.[5]

So who provides the goods?

I should add a final note on the *provision* of merit goods. It is a common error amongst students that these goods are always produced by some form of national or municipal

---

4 Credo set in stone by *James Dean*.
5 I fully realise that my use of condoms as an example in a textbook aimed at 16 – 19 year olds will cause a few hiccups of shock in some cultures but I offer no apologies. The issue of HIV/AIDS has moved far beyond the niceties and confines of morality and/or cultural relativity (ask your TOK teacher) and boils down to saving lives. The 0.06 mm of latex might make the difference between life and death. Talk about marginal benefits….

government. Not so. You must distinguish between 'provision' and 'production'. In most countries merit goods are provided for (here 'paid for') by government yet many countries have extensive tendering systems whereby private firms compete for government contracts. Tax money indeed *provides* the good – but it is *produced* by the lowest bidder.

In any case, all societies must deal with some sort of public provision of goods and thus with an element of planning in the economic system. Even if the society is based on an American model, where there is a far larger element of personal payment into health and pension funds than in Europe one must have a degree of public health and social security. Again, politics must decide this at the behest of economics and other social sciences. Note that it is virtually impossible to define which goods are actually merit goods on the basis of economic reasoning alone. The societal attitude that certain goods are desirable is in fact a political decision. Not all societies consider ballet, evening courses in carpentry or martial arts to be goods which should be provided by public monies. The political decisions herein help to characterise merit goods and help to fulfil the objective of *redistributing incomes*, since poorer groups gain access to goods often provided by progressive income taxes.[6]

Government policies for merit goods

Education is considered one of the most vital merit goods, as it increases not only de facto output (by increasing the efficiency of labour) but also future *possible* output (by providing the basis for scientific research, development and innovation). This is a good which is grossly undervalued in many richer countries with tax-funded schooling, probably since we have become so used to not paying the actual cost of provision. I had the nasty habit of telling my Swedish public school students that only one out of 20 of them would be sitting in class if they had to pay the actual cost of an IB education. The thought-provoking thing is that it is not so much a *lack of ability* to pay, rather than a *lack of propensity* that would preclude the other 19 from joining classes. We are disinclined to pay for something when we cannot see the benefits of consumption. Putting it harshly, as I am prone to do, one could say that the inability of the average 16-year old to assess the value of education will harm not only the individual student in the long run, but also society in the guise of geriatric economic teachers who look to well-padded

---

6 A progressive income tax means that the proportion of tax (i.e. average percentage) increases as income rises. Higher income = higher average percentage tax paid.

societal coffers for help to the bathroom in old age. This social service is provided by tax revenue. Tax revenue cometh from economic activity – which is enhanced by education![7]
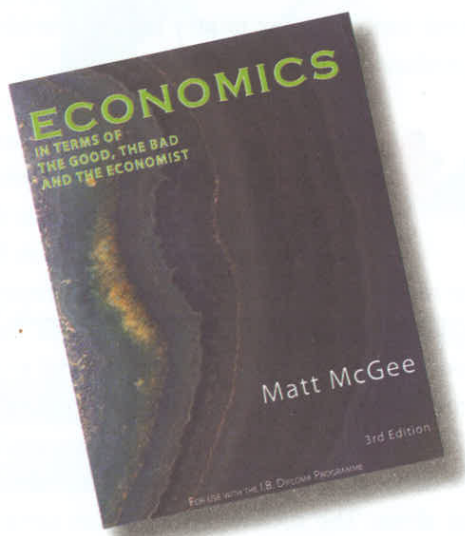
"Buy me! Live better!"

Basically, governments adopt some of the following measures to increase wider availability and consumption of merit goods:

- **Increasing demand for merit goods**: Many cities use advertising campaigns to get more people on bicycles in order to lower inner-city congestion and pollution. Municipalities send households information brochures on recycling centres. In Mexico I was staggered by the numerous murals (= wall paintings) in rural areas sponsored by Federal Government to get people to use mosquito nets against dengue fever.[8] The list is long and you will have little difficulty in finding numerous examples simply by looking around you on your way to school. In Figure 18.2, demand increases from MPB_0 to MPB_1 – closer to the societally optimal point of output at $Q_{soc\ opt}$.

---

7   We're back to self-interest as a motivator. Maybe this should be my strongest incentive to be as good a teacher as possible; the thought that someday the nurse who helps me to the toilet in my old age will be provided by tax revenue generated by my own students. Please make sure the nurse is equipped with a sense of humour and a cigar-lighter. Don't worry about the nurse's age, gender or looks. I'll be far beyond remembering why this would be of any possible interest to me.

8   Dengue fever (pronounced "*den*-gee") is caused by a mosquito borne virus similar in geographical spread to malaria. While seldom fatal, the aches, pains and fevers associated with dengue are quite severe. My father had both dengue and malaria – and claimed that dengue made malaria seem "…like a boy scout picnic".

- Increasing supply of merit goods: Governments can either provide the good – education and health care are common examples – or subsidise the good. The effect of a subsidy is shown in Figure 18.2 as an increase in supply from MSC to $MSC_{+subsidy}$ – again moving consumption towards the societally optimal point of $Q_{soc\ opt}$. There are numerious examples of goods subsidised by governments – often without consumers knowing it. *Dental care* for the under-16s, basic foodstuffs such as *rice* or *milk*, and local *sporting events* are common examples of subsidised goods. Yet there are a number of goods which increasingly come under the scrutiny of government officials 'subsidy microscope'. Such goods are often subject to *purchasing subsidies* (often by lowering value-added taxes or other indirect taxes). For example, a number of countries offer lower road taxes for owners of hybrid cars which run on both petrol and batteries. In a similar vein, some governments offer tax deductions for house owners who install solar panels and/or additional insulation.

A number of questions arise in evaluating government actions to increase consumption of merit goods:

- Many question whether the government is in fact the most efficient provider – could not private companies provide the good more efficiently?

- Which goods should be subsidised or provided by taxpayer monies? Where does government draw the line if government provides?

- Subsidies are a cost to government and therefore increase the tax burden on households and firms. This can have a negative effect on consumer spending and firms' investment.

- It is almost impossible to estimate the *future* benefits of an increase in *present* consumption of merit goods! This makes it very hard to arrive at a "correct" level of provision.

- Many people react to the "nanny state" mentality, e.g. a government which imposes an overprotective and even stifling net of regulations and market intervention measures.

It bears pointing out that the costs to government in the form of funding the subsidies are considered to be less than the possible benefits – otherwise the subsidy would not exist. In other words, there is a **net social benefit** in increasing production/consumption of the good, the problem being that the private sector would not take care of the additional output without the incentive and cost-cutting effects of subsidies.

Summary & revision

1. **Positive externalities in production** are when the marginal private costs are higher then the marginal social costs (MPC > MSC). In terms of resource allocation, the good is underprovided. Fire breaks protecting individual properties is an example of a good with positive externalities in production.

2. **Positive externalities in consumption** exist when an individual's consumption of a good has positive effects on non-users/consumers. Thus the marginal social benefits are greater than the marginal private benefits (MSB > MPB) and the good is under-consumed. Health care, education and pension schemes are examples of goods with positive externalities in consumption.

3. **Merit goods** are goods which have benefits to the user in the future and also positive spillover effects in consumption.

4. Since merit goods are **under-consumed** in terms of societal optimum (e.g. where MSC = MSB), government policies aim to increase demand and/or increase supply of merit goods:

    a. **Increasing demand:** positive advertising campaigns for bicycles, incentives for households in developing countries to send children to school, and information to households about recycling are all possible government measures.

    b. **Increasing supply:** government or the municipality can supply the good at break-even price or even at a loss, subsidise the good, and offer various forms of incentives for firms to provide more of the goods.

19. Lack of Public Goods

Key concepts:

- Private and public goods – classic division of goods in economics
- Free rider problem
- Provision of public goods

Private and public goods – classic division of goods in economics

Well, it's getting late and I dearly want to head back home for steak, beer, TV news and lots of questions from the Very Small Australian Female I am married to. I could walk but hey, this is Jakarta, so I'll put on my manta ray cowboy boot and call my driver. Here's where the concept of **public goods** comes in: I will be using the **public roads** to drive home on – guided by **streetlights** and **road signs** galore. I will studiously avoid any and all contact with the Indonesian **police**. The **blue** examples above are goods which would all be **under-provided** or not provided at all by a competitive market.[1]

Imagine if, in driving home, I had to stop and pay a fee for road use (which is actually becoming increasingly common) and then had to put a few coins into coin-operated street-lights and coin-operated road-signs along the way. (I was about to say something about having to pay the police on the way – but

1   Once again, note that economists use the term 'goods' as an umbrella; it covers tangible goods such as guns and butter but also services such as health care and banking. Public and merit goods are, in fact, often services.

of course that is exactly what one does here in Indonesia.) The common denominator here is that the thought of paying for one's actual 'unit' use of the goods outlined here borders on the preposterous – not to mention impossible.

Roads, street-lights, road signs, police service and public broadcasting are all examples of **public goods**; they are 'publicly' available (as opposed to 'pay per usage'). Be careful in your use of this term! It is one of many unfortunate examples of how economists have corrupted the common use of a word and created a very definite subject-specific term. When economists use the term public goods they are referring to something quite specific; a good which by its very nature is impossible to charge for on a user-pays basis. A most treasured example of public goods are *ideas* – pasteurising milk for example, using coal/gas/uranium as a source of energy, and all forms of advances in farming technology – noting, of course, that the development of the horse- or oxen-drawn plough didn't run into the highly debated current practice of patenting genetically modified organisms. The two main traits – pre-requisites, actually – of a public good are **non-excludability** and **non-rivalry**.

Figure 19.1 – classic economic division of goods

Division of goods	Exclusion of non-payers (excludability) – is it possible to charge users separately and based on the quantity consumed?		
		YES	NO
Competition in consumption (there is rivalry or diminishability in consumption) – does my use diminish others' use?	YES	Private goods: food, cars, houses, clothing	Common access resources: natural environment, open range grazing land, fish stocks in international waters
	NO	'Club goods': clubs, private schools, public parks, TV broadcasts	Public good: police force, lighthouses, air traffic control, street lights, fireworks displays, military defence, earthquake or tornado warning systems, flood banks

Non-rivalry

By putting on my manta ray boots to go up to Toni's, I *hindered others* from consumption of the good, which is an example of *rivalry* – my use of my boots bars others from use. My boots, car and MP3 player are examples of rivalry in consumption, since my use decreases the availability of the good for others. A public good, by contrast, does not have this element since the good can be used simultaneously by other people. My use of the road, road signs, police force, and streetlights does not reduce available consumption to others. Thus public goods are subject to (varying degrees of) *non-rivalry*, meaning that one person's use does not diminish the availability of the good for others. *Get off of my cloud* by the *Rolling Stones* is actually far more meaningful than 'Get out from under my streetlight'![2]

Non-excludability

What about hindering someone from using the good in the *first* place? Yes, this is actually the definition of private goods; once the good has been provided, the provider must be able to *exclude* you from use – in order to be able to charge you. Thus the provider can exclude non-payers from use, just like the boots, car and MP3 player. This highlights precisely the problem with goods such as roads, road signs, and street lights; once they have been provided they can be used by anyone – payers and non-payers alike! This means that there is an element of *non-excludability* built into some goods, and this is the second pre-requisite of a pure public good; once provided, one cannot exclude anyone from using it.

Is it non-rivalrous even in daylight?

2   But there's no way I'm not going to tell the following little story! My friend Toni lives in a 'gated community' – e.g. an area behind gates where you have to show ID to get in. After some heated discussion in the community about who was paying for the streetlights or not, the community leaders assigned streetlights! Anyone who didn't pay their bill found themselves on a very dark street. Another favourite example in economics bites the dust…

Free rider problem

This is also known as the **'free rider'** problem and should be well-known amongst students who have ever done a group paper where three students did all the work and the fourth was a useless slacker. The term 'free rider' is actually taken from public transport, where the provision of a bus service would be 'free' for those who didn't pay the fare yet provided for by those who did. Hence 'free rider'.

> ### Definition: 'Public goods'
> A **public good** is distinguished by high positive externalities but also non-rivalry (my use does not diminish your use) and non-excludability (non-payers cannot be excluded once provided) in provision. It is very difficult to charge individual users and is thus often provided by public monies.

Provision of public goods

There is a very noticeable common denominator amongst the examples of public goods above, namely that they are in fact considered of such importance that they are supplied! There are in other words goods which cannot really be charged for – much less limited to users only – yet are provided for in virtually every modern society. Why? How? The answer to the first question is by now self-evident; any good which is considered to be of such obvious gain to society – roads, lighthouses, drainage systems, water filtration, electricity grids, military and civil defence, air traffic control, police force, judiciary system, etc – will ultimately be both desired and supplied somehow. The remaining question, 'How?' must enable potential suppliers to be able to provide the good economically in spite of not being able to charge for its use. This would exclude the private (free market) sector immediately. The answer is **taxation** and **public sector provision** of the good. Everybody pays since everybody benefits. End of story. As everyone is reaping the benefits of consumption it is only fair that everyone 'pull a straw from the stack' and help pay for it.

Free rider?

On a rhetorical note, one must include a few weighty objections to the concept of a *pure* public good. Anyone who's been stuck in gridlock in the outer city access roads is well aware that your use does, in fact, diminish my use – something that any number of drivers in Athens will gladly attest to! The same goes for getting hold of a police officer to investigate a break-in to one's garage at the same time the community is swamped by either football hooligans, a rock festival or anti-globalisation demonstrators. It is also evident that increasingly sophisticated pay-per-use systems allow communities to charge for the use of inner-city roads and access routes. In truth, there are very few examples of goods which would fulfil the criterion of non-rivalry and non-excludability to 100%.

Public goods also come with large positive externalities, yet with the troublesome attachment of being very difficult to charge users for. As outlined earlier, few goods could be considered to be pure public goods. Roads can be toll-roads (= 'pay per use') and police services can be unavailable at peak times which means that the non-rivalry and non-excludability criteria are not met. Examples of *pure* public goods would perhaps be defence, air traffic control and lighthouses.

Since the private benefits would be very small (or non-existent) compared with the public benefits, competitive markets would fail quite massively here. The good would simply not be provided in many cases and the social welfare losses would be very large. This is probably why the debate on whether public goods should be supplied by public monies is much more subdued. In fact, virtually all countries have a degree of government supply of public goods. It seems that there is broad agreement that lighthouses come with sizable social benefits – such as not having an oil tanker grounded on a holiday beach or the nesting ground of a seal population. And seriously, can

you see air traffic control demanding that an off-course airplane outside of Washington pay a fee in order to be brought in safely?

What is potentially contentious is the *quantity* to provide since there is no market mechanism to gauge demand. The market fails to achieve optimal resource allocation and so society must make up for it. This is another example of how the planning element cannot ever be completely discarded – all economies must have a mixture of private and public it seems. Yet, once again, it is quite common for public goods to be supplied (as in *produced*) by private firms which have been contracted by government. Municipal government pays private mountaineers to dynamite overhanging snow-shelves in the Vorarlberg alps in Austria to avoid avalanches while on the other side of the earth the central government in Laos – one of the world's last centrally planned communist states – pays a private Swedish company to build 600 km of road, Highway 8[3].

Summary & revision

1. **Public goods** have two basic characteristics other than high positive externalities:

    a. *Non-excludability* – once the good has been provided it is impossible to prevent non-payers from consuming the good

    b. *Non-rivalry* – on person's consumption does not diminish (or rival) another person's consumption

2. **Free rider problem** means that it is possible to consume a good or resource without paying for it. This is a problem common to public goods and thus precludes provision by a free market.

3. **Examples of public goods** are police force, lighthouses, air traffic control, street lights, fireworks displays, military defence, earthquake or tornado warning systems and flood banks

4. The **solution to market failure** in the case of public goods is to provide the goods via taxpayers' money.

---

3   Apparently Japanese firms have been contracted to build the bridges across some 15 rivers. Unfortunately, while many stretches of road have been completed, **none** of the bridges have been built! This might be a good example of when central planning/provision might be superior to private.

20. Common Access Resources and Sustainability

Key concepts:

- Common access resources
- Sustainability
- Lack of price mechanism – market failure again
- Another trade off – poverty and land exploitation
- Possible government responses to sustainability problems
- The international dimension of the sustainability problem

Common access resources

In Chapter 19 (Figure 19.1) a public good was defined as non-rivalrous and non-excludable. When a good is indeed 'rivalrous' in consumption but non-excludable (impossible to charge users separately), we speak of common access resources. In other words, we are looking at resources such as free-graze land, open access rivers and free range forests where individual users cannot be charged for use but where indeed an individual's use diminishes the use of others. The basic problem is once again the lack of property rights; since there is no clear owner and all have access to the resources, there is little or no incentive to limit use and/or take care of existing resources. Adding to this increasing populations and the burden created on available land, it is no surprise that the results are deforestation, soil erosion, pollution of rivers and lakes and decimation of wildlife.

> **Definition: 'Common access resources'**
> Resources which by their nature are rivalrous but non-excludable are know as **common access resources**. Forests, lakes, rivers and oceans are examples of common access resources.

Perhaps the scariest example – and one even very optimistic observers note – is the on-going over-fishing in seas and rivers. Since one fifth of the world's population is dependent on fish as a main source of protein, it is hardly surprising that depletion rates are such that an estimated 70% of the world's fish species are either fully exploited (e.g. on the edge of sustainable fishing) or depleted.[1] Rising populations and ever improved methods used by fishing fleets are the root cause of the severe overfishing taking place. And while several attempts have been made to put together internationally binding conventions and quota systems, one must face the facts that none have seen any success so far.
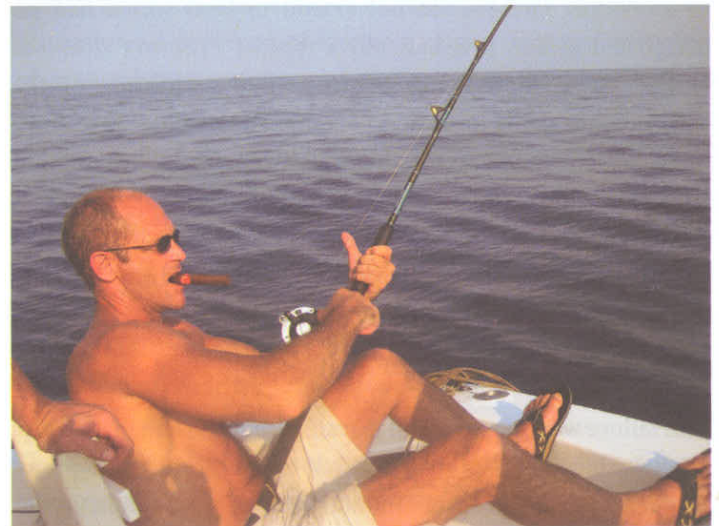
Catch 'em before we run out!

...unless I catch you first!

---

1 http://www.un.org/events/tenstories/06/story.asp?storyID=800

161
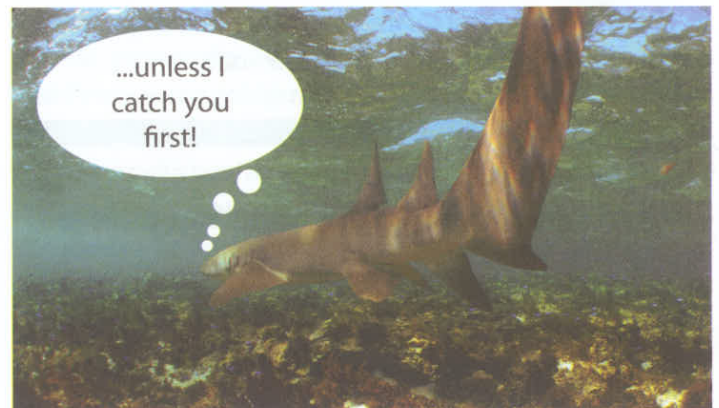
A Little Depth

The tragedy of the commons

Every once in a while, history and the coining of a phrase has applicability and usefulness for a wider audience. Such a phrase is the **tragedy of the commons**, which refers to land set aside by villagers in the 17th century which was to be owned by everyone, i.e. community-owned pastures. This common (where 'commons' comes from) land was freely available to all who wished to use it for grazing cattle. The indistinct property rights ultimately resulted in a 'Do it to others before they do it to you' attitude.

Villagers would try to avoid over-grazing on their (privately) held land in order to keep it as high-yielding as possible. This, of course, did not extend to the commons. The self-interest of individual farmers did not extend to land which had no designated owner. It is basically a type of *Prisoners' dilemma* where putting another cow on the pasture is 'cheating' on the total possible gains for others – compliance would be to limit one's use to increase everyone's overall gain. Thus the common land was subject to hard use and ultimately suffered from continuous overuse – abuse even. Overgrazing, poor drainage, lack of fencing all contributed to lower the carrying capacity of the land to the point where it became virtually useless; a squandered resource.

This failure was due to lack of private incentives and ownership. As many other examples have arisen over time, any case where inadequate or non-existent maintenance of public domain has been referred to by economists as a 'tragedy of the commons'. Over-fishing in international waters and overuse of antibiotics (creating resistant strains of diseases) are good examples of how this tragedy is reprised on a much larger scale.

Lack of price mechanism – market failure again

When I bring the issue of resources being used in such a manner that damages are not paid by the user, I ask my students if their parents have ever rented a car. Hands go up. Then I ask if by any chance a dad (it's *always* the dad!) ever said something to the tune of "Let's let it rip and see what this baby is capable of!" and then rammed through the gears and drove over bumpy roads much too fast. A few sheepish hands go up. I then explain that this is simply because the cost of repairing the suspension and gear box has been *put forward* on to a *future* driver. It's really no different when one looks at the use of limited – and non-renewable – resources; it's time inconsistent behaviour. Firms' and households' use of coal, forests, oil, water…etc will inflict costs on future generations.

The basic problem of common access resources is one of hazy prices in using limited resources – users do not take into account the marginal social costs of over-use. Use of limited natural resources which lack clear property rights is in fact an example of negative externalities in both production and consumption. (See Figures 17.3 and 17.4 in Chapter 17).

- **Externalities in production**: When power companies destroy land to mine coal, destroy hectares of forest to build dams and cause acid rain by burning oil for electricity, the external costs are considerable. Third parties such as local villages living with the particulate pollution of coal mining, soil erosion and destruction of rivers bear costs unaccounted for. In economic shorthand, the MSC > MPC.

- **Externalities in consumption**: Using natural resources at a rate beyond a sustainable level give individual consumers benefits at the expense of future consumers. Overfishing, over-grazing and excessive logging in rain forests will cause fewer possible benefits in consumption for third parties. This negative externality in consumption means that the MSB < MPB.

Sustainability

A far greater threat to overall well-being is the possibility that environmental externalities ultimately make it evermore difficult to sustain output increases over time. In the short run, the production of goods will cause environmental damage such as air-borne pollution and toxicity in rivers and will often use non-renewable resources. The basic issue is not *whether* growth is detrimental to the environment and uses limited resources, but rather to *what degree* this is sustainable in the long run. A country with a real growth rate of 2% per year will have doubled its income in 35 years. This could have serious implications for total use of non-renewable resources such as oil and coal. Most economists would agree that a great deal of output uses resources which are irreplaceable in the short run and that the

ability to increase long run output is contingent upon improved technology and production methods.

Consider energy sources as one example of the short term versus the long term. Any form of energy using fossil fuels (oil, natural gas, coal) will have severe repercussions on the environment. Burning fossil fuels means increased soot and particulates, carbon dioxide/monoxide which contributes to the *greenhouse effect* and thus global warming. Nuclear energy on the other hand, is possibly the cleanest form of economically viable energy forms in existence at present – in the short run. The long run hazards of dealing with the nuclear waste which have lethal half-lives of thousands of years are inestimable – and put quite a burden on generations as yet unborn. The issue is in fact far more complex when viewed from an economic perspective, since it is a fallacy to believe that the ideal quantity of pollution would be zero. This perspective fails to consider that this 'Utopian' state of being will have severe opportunity costs in itself, such as far less available energy at higher cost.

Another trade off – poverty and land exploitation

Let's apply the concept of **inter-temporal** ('between two time periods') opportunity costs to the question of environmental issues. If society desires present output and growth, then it will in all likelihood see pollution, soil degradation and deforestation as a cost. However, in focusing on future availability of clean air, ample soil and forests, then society will forgo an amount of present output. Play now and pay later – or vice versa?! Quite frankly, this issue has been answered in a rather negative way in many of the rich industrialised countries; forests, minerals, land and oil were severely depleted during the initial phases of development throughout the 18th to 20th centuries. This in fact created a great deal of the present wealth in developed countries and highlights the conflict of interests today in less developed countries (LDCs); how far should these countries go in limiting their use of natural resources in order to sustain future levels of output – knowing full well that this limits present growth and present levels of living standards? There are also arguments supporting present use of resources in order to attain technological advances which would enable us to use fewer resources in the future.

Pessimistically speaking...

The basic definition of sustainable development is that present use of resources in satisfying needs in an economy should not lessen or limit future generations' use of resources in satisfying needs.[2] This definition is slippery at best. Any action taken by a developing society today *will* have an impact on future output and *will* involve using limited non-renewable resources such as minerals and oil. Furthermore, it is a common (and faulty) assumption that when the production of a good impacts negatively on the environment, the production should be banned. This is not an economic analysis since many of the goods whose production pollutes will also have benefits to society – cars for example. Economics aims to assess the **net** marginal benefits, which entails taking into account additional external costs and benefits in setting output and price. Basically, economists provide a way of looking at socially acceptable levels of pollution. In this harshest of lights, it is clear that it is impossible not to adversely affect future generations' use of non-renewables to some extent.

Optimistically speaking...

On the other hand, improved technology and production methods have enabled more developed countries (MDCs) to substantially increase output whilst lowering the use of natural resources. In many cases, such as oil, copper and coal available 'years left at present consumption-rates' of these finite natural resources have actually increased in spite of population increases and depletion of known reserves. This is a reflection of how relative scarcity and the price mechanism leads to productivity increases and the search for substitutes; the cost of telecommunications has plummeted due to increasing use of computers and also the replacement of copper wires with optic cables made of fibreglass – that is to say, sand! Therefore, while the pure physical availability of many resources is diminishing, one can argue that improved technology/efficiency in production means that we are providing the next generation with a better production base than we ourselves have.

The debate on development, growth and sustainability is fierce and multifaceted. Many environmental groups claim that present consumption and production is impossible in the long run and even that our world will not survive if developing countries were to follow the same growth and consumption path as present-day MDCs. 'Imagine the same ratio of cars and Freon-using refrigerators in China as in the USA!' is a common argument. Countering this is the above argument on how production in MDCs is increasingly efficient and the environmental costs progressively lower. Therefore increasing the income in developing countries will enable a transfer of technology from richer countries and ever improving output-to-pollution ratios.

2 The concept was initially used in the aforementioned (United Nations) *Brundtland Report* from 1987.

Environmental Kuznets curve

The complexity in trying to analyse the costs and benefits involved in growth and development has received a good deal of attention. Recent studies and increasingly clear data indicate correlation between economic growth and environmental degradation, primarily airborne pollutants.[3] Note that I did not use 'positive' or 'negative' correlation. The reason is simple; at low levels of income an increase in national income will be associated with increasing environmental costs (positive correlation between growth and pollution), while at some point along the growth curve increasing income will show improvements in the environment (negative correlation).[4] Figure 20.1 shows how increasing per capita GDP is correlated to a decrease in environmental degradation.

Pollution per money unit of GDP rises at first and then peters off as national incomes rise. This is primarily noticeable in emissions of particulates and sulphur dioxide (SO_2).

Figure. 20.1 Environmental Kuznets curve; growth and environmental impact

At low levels of income the amount of pollution is relatively nominal and increases as incomes rise. However, as per capita incomes rises the relative intensity of pollution (= pollution per unit of output) in the economy falls, as technological improvements take place and environmental standards are raised. The reasoning is graspable at an intuitive level; when people become wealthier they are more willing to pay for a cleaner and healthier environment. In actual fact, they would demand it, and increased income provides the resources necessary to clean up the environment. Unfortunately, the curve in Figure 20.1 has not been shown to apply to the greenhouse gas, CO_2, which solidifies this as one of the major environmental issues to address.

POP QUIZ 20.1

Externalities and Sustainability

1. Using diagrams, explain how economic activity can negatively impact on society. Use both consumption activities and production activities in your examples.

2. What would be the main difficulties in assessing external costs and benefits?

3. Given that all economies are to some degree reliant on finite resources – for example oil – explain how it is possible for the present generation to leave the next at least as well off in terms of being able to meet their own needs.

Possible government responses to sustainability problems

There are several options available to governments trying to address sustainability issues in production and consumption. In **Chapter 17** we looked at taxes and the 'polluter pays' principle, e.g. tradable permits (cap and trade permits). Other solutions in the case of common access resources are bans and strict regulation – for example a ten year moratorium (= freeze) on Atlantic cod fishing, strict catch limits of Baltic salmon or logging bans on Borneo. However, the issue here that is so hard for governments to deal with is simply that a great deal of our common access resources are subject to agreements between

---

3 See, for example, the '**Environmental Sustainability Index**' put together by the World Economic Forum's Global Leaders for Tomorrow Environment Task Force, The *Yale Center* for Environmental Law and Policy, and the *Columbia University Center* for International Earth Science Information Network (CIESIN); http://www.ciesin.org

4 *Alan Kreuger* and *Gene Grossman* of Princeton University estimate that the turning point (in terms of GDP/cap) is around USD5,000 – roughly where the Czechs are now. By the time per capita income is at USD 8,000, all environment indicators show improvement. See *Economic Growth and the Environment, National Bureau of Economic Research*, NBER working paper W4634, February 1994 See also *World Bank, Policy*, research working papers 1992; no. WPS 904. World development report by *Shafik*.

countries. It is virtually impossible for a single nation or small group of nations to deal with air pollution and international fish quotas since these transcend (= go beyond) geographical boundaries.

The international dimension of the sustainability problem

It should come as no surprise that most of the possible solutions above involve a great many problems of implementation due to the simple fact that externalities know no borders. Drug use, for example, is an international problem since user countries are often not the same as producer countries, making clear that any real solution will involve attacking not only use but supply. In pollution, there are massive difficulties in getting a country to pay for externalities inflicted upon the domestic economy of another. In all, these issues require extensive arrangements in cooperation which is why there are an increasing number of international agencies involved in information-sharing, monitoring and assessment. In dealing with negative externalities on a global basis, there are a number of difficulties involved. Here are the main issues:

Accountability

Clearly it would be rather difficult to assign 'pollution portions' to different countries. A country which suffers from acid rain resulting from emission in another country will have a most difficult burden of proof in showing who the culprit is. Just how much of the acid rain falling in Belgium is the result of factories in France?

Cost estimation

Once again, one of the basic problems in ascribing costs to negative externalities is setting appropriate costs. Most of us would react if we were asked to put a 'price' on a loved one, and correctly pricing the environment runs into similar subjective and normative problems. How would one estimate the loss of biodiversity in lakes in terms of money?

Method

There is great debate as to which methods are the most efficient. While the US argues that market-based systems such as pollution permits are the most efficient way to lower overall pollution levels, there has been widespread criticism from environmental groups that neither taxes nor permits squarely address the root of the problem. It is often pointed out that (due

to time lags in cause and effect) we have not yet seen the total results of emissions, for example, that even if every industry in the world shut down tomorrow, global warming would continue for a number of years.

Trade-offs

The OECD estimates that CO_2 emissions – accounting for 60% of greenhouse gases – will increase by roughly 33% by the year 2020. This could have severe repercussions when countries try to strike a balance between growth and environmental impact, as the opportunity cost of lowering CO_2 will mean lower growth rates.

Loss of competitive edge

Stricter legislation on the environment will drive up production costs for firms. There will be a clear incentive for countries to accept laxer environmental legislation in order to remain internationally competitive and keep both growth high and unemployment low.[5]

Multilateral environmental agreements (MEAs)

The above notwithstanding, the fact remains that it is in everyone's interest to lower environmental damage in the world. Clearly, any long term solution must involve all the contributing and affected countries and have clear mutual benefits. To this end a number of organisations, often under the collective umbrella of the United Nations (UN), have worked to create international environment treaties and agreements. An increasing number of **multilateral environmental agreements** (multilateral meaning that a number of countries are involved rather than an agreement between two countries), so-called **MEAs**, have been implemented in the past decade and more are subject to ongoing negotiations. MEAs cover topics ranging from trade in endangered species and hardwood to regulations on the use of chlorofluorocarbons (CFCs) and emission caps on CO_2.

5 There has even been some speculation that countries would even compete in having the poorest environmental standards in order to attract foreign manufacturing companies. This is the 'race to the bottom' effect which many environmentalists have long predicted. As yet, there is little – if any – evidence of this happening. More on this in Section 4.2.

MICROECONOMICS

The Montreal Protocol

In 1987 a total of 41 countries (12 acting under the EEC – now the EU) signed an agreement to severely limit and phase out the use of substances that deplete the ozone layer. These original signatory countries accounted together for over 80% of total ozone depleting emissions when the agreement came into effect in 1990. Since then, other countries have joined, whereby 185 have now ratified the agreement. The primary method has been to set strict (and legally binding) caps on emissions of ozone depleting substances such as chlorofluorocarbons used in refrigeration and aerosols. These chemicals have been phased out and replaced with ozone friendly substances. According to recent research, the agreement has been a success, with levels of harmful chemicals falling and ozone levels actually slowly recovering.[6]

The Kyoto Protocol

The UNFCCC (see above) provided the basis for the 1997 agreement on limiting greenhouse gases (GHG). The Kyoto treaty was partially modelled on the success of the Montreal Protocol, but has been something of a failure so far. The scheme called for industrialised countries to reduce their emissions of GHG by an average of 5.2% from the 1990 level over the period 2008 – 2012, using a mixture of emissions trading and mechanisms to earn 'emission credits'. The Kyoto Protocol came into force when *ratified* (= formally signed, a binding agreement) by 55 countries which together accounted for no less than **55% of global GHG** emissions in 1990. During the initial phase, developing countries were exempt from emission curbs.

These signatory countries accounted for 44.2% of CO_2 emissions, leaving 10.8 percentage points left before the agreement could come into effect. Only five industrialised countries initially refused to ratify the agreement, amongst others the USA and Russia, who together account for more than a third of global GHG emissions. By February 2005, after Russian ratification, the necessary 55% of "GHG contributions" had been reached and the Kyoto Protocol came into effect.[7] By 2011, 192 states (comprising circa 63% of GHG emissions) had ratified the treaty. Notably, the world's largest emitter of GHG, USA, has still refused to ratify the agreement claiming that it would be too costly for US industries and that notable contributors to global emissions (China and India) are largely exempt from the agreement.

The Kyoto debate

The *pro et con* debate has been high-pitched and very politicised. *Proponents* of the Kyoto Protocol refer to the proposal as a first step of many and the only truly global attempt to diminish possibly the greatest threat to our world. Supporters of the treaty point out that the USA accounts for 4% of the world population but emits 16% of the GHG (2005 values), which would make the treaty close to futile even if all other countries were to ratify. It is pointed out that since the present industrialised nations have caused over 80% of present GHG and are the only nations which have the economic and technological ability to clean up their act, the brunt of the burden of GHG cuts should go to them rather than spreading responsibility 'evenly' by including developing countries which are not the culprits. Environmental groups have frequently stressed how strong business links and lobby groups in America have been instrumental in shaping US environmental policy – Greenpeace went so far as to say that president Bush was 'really being led by ExxonMobil.'[8]

Opponents of the protocol point to the heavy economic burden that implementation would bring to the USA, the world's largest economy; a reduction in emissions of an initial 7% resulting in the long run of some 35 – 40% overall reductions from 1990 levels during 2008 to 2012 at enormous costs in terms of global GDP – the OECD estimates the total cost to be 2% of OECD countries' GDP by 2050. In addition to this, there seems to be little evidence that world GHG emissions are able to be kept within Kyoto limits for the foreseeable future. There is also an argument that the Kyoto Protocol will in effect only delay a rise in global temperature by a meagre 6 years during the coming century – the 0.15° C rise in temperature estimated to result in the year 2100 will instead arrive in 2106 if Kyoto is implemented and carry a price tag of a 2 – 4% reduction in GDP.[9] US officials are quick to point out that most developing nations – primarily China and India – were to a large extent exempt from the Kyoto protocol, but are quickly catching up with industrialised countries in terms of GHG emissions, which would mean that the most efficient producers would be harnessed while the most inefficient would be exempt from the treaty. Finally, one could mention that research is inconclusive

6 http://www.unep.org/ozone/pdf/execsumm-sap2002.pdf. and
   http://worldbank.org/montrealprotocol

7 http://UNFCCCc.int/kyoto

8 http://www.greenpeace.org.au/climate/government/kyoto/
   index.html

9 *Lomborg*, pages 302 – 305

as to whether on-going global warming is *primarily* man-made or part of a larger general climate pattern.[10]

The future of the Kyoto Protocol

The Kyoto agreement was a first step in an on-going process aimed at reducing GHG. It expires in 2012 and in November 2011 signatory nations met in Durban to assess the results and also decide the next round of GHG emission caps. A main issue was that Canada, Japan and Russia professed little interest in signing a new treaty round of GHG reductions. The US and China seem to prefer individual action rather than a global treaty. Another issue is the extent to which developing nations will be bound by the same rules of capping as developed nations.

Summary & revision

1. **Common access resources** are rivalrous but non-excludable in consumption. Forests, lakes, air and oceans are examples.

2. Over-use and destruction of common access resources is a form of market failure since there are **externalities** both in **production** and **consumption**.

3. LDCs face the **trade-off problem** in use of natural resources. Poverty levels and growing populations put huge burdens on land use and severely limit future use of limited – and often non-renewable – resources.

4. Government responses include **taxation** of polluting industries; **bans** on land and water use; **tradable permits**; and **quota systems** for fishing.

5. **Key problems** in dealing with market failures arising from over-use of common access resources is that of **accountability** (who is the 'culprit'); **estimation of costs**; 'best' **method to correct** the failure; and the fact that many of the external effects transcend national borders which necessitates **international cooperation**.

MICROECONOMICS

---

10 You are going to have to read up on this on your own. The more I read on the subject of global warming, the more confused I get.

21. Asymmetric Information and Imperfect Competition

Key concepts: HL extensions

- Information asymmetry and market failure
- Imperfect competition and market failure

Ponder the following two examples taken from the Life and Times of McGee and then read on. I tell you what I tell all my students; my examples *sound* flippant and off-the-cuff but in fact are chosen for a reason!

1. After rolling my VW Jetta four times doing 140 kms/h at 3000 meters above sea level on the road between Mexico City and Cuernavaca, I decided to move up-market and get a used Audi A6 Quattro with 250 horsepower. I got a really good price at the dealer and drove off in self-contentment and pride. Within six months I had spent an additional 25% of the purchase price on repairs. After a year I was prepared to *pay* someone to take it off my hands.

2. Lady Bell just bought me my 50th birthday watch! A lovely little diving number made by Steinhart and waterproof to 1,000 metres. After hours of searching on Internet sites, she found a seller in Bratislava, Slovakia who beat all the other sellers in terms of price, delivery time and warrantees.

3. In Jakarta where I now live, if one wants a good bottle of wine one goes to a very limited selection of shops since Indonesia is predominantly Muslim and the government has strict limitations on sales of alcohol. When I wanted a good bottle of Rioja to surprise the Very Small Australian Female (Lady Bell, my wife) on Valentine's Day, I almost had to sell a kidney to pay for it.

In one particular case I would put forward that I paid 'too much'. Have a glass of Rio…milk and a cookie while you ponder about what 'too much' means and *why* I, the hard-nosed economist, got the wrong end of the stick.

Information asymmetry and market failure

Yeah, you worked it out didn't you?! The car salesmen knew about the bad tires and faulty transmission system. I didn't. (Look ahead to Chapter 27 and the assumption of '…perfect knowledge and information…' in our perfectly competitive market model.) If two parties in an economic transaction have complete information about quality and price of alternatives, then both will make a decision that maximizes marginal utility. If then the transaction takes place, neither can have been better off in any other transaction and the market clearing price and quantity result in perfect resource allocation.

Yes, you can hear the mermaids and unicorns dancing in the background – e.g. this is not particularly realistic but fairy-tale economics. Could one in fact *ever* have 'perfect' information in purchasing a good?! In the case with my car purchase, the information was asymmetrical (= uneven) and – I hasten to add! – in the seller's favour. He could set the price higher than I would have been prepared to pay had I had perfect knowledge and information. If *all* transactions in society were asymmetrical then all consumers would be paying more than their marginal private benefit. Markets fail.

In the case of the wine, well, I really didn't have a choice and bought it so it means I considered it worth it and I added to my marginal utility by a fraction. Now, in the case of my new watch, I did rather well – e.g. I maximised my marginal utility…at least I *think* I did. Let me live with my delusions. In any case, this is just an example to point out that modern technology and the Internet have a very strong *enabling* function for consumers. We are able to search for information in a few minutes and this empowers us in our comparison shopping. My access to more complete information tips the scales a bit and decreases the information asymmetry so often in favour of sellers.

Happy birthday!

What about asymmetry in favour of consumers?

While not part of the syllabus, I cannot help but pointing out that there are some real life examples where consumers have information that sellers lack. The classic example is when one buys health insurance or full coverage car insurance.

Information asymmetry in the case of life insurance is an example of *adverse selection* – I would most assuredly NOT tell the insurance company about all the injuries I have sustained during 30 years of knock-down karate! This would cause them to charge me more…or even cancel my insurance. I do not disclose this information and of course many others would do the same. This is the 'adverse' (= contrary) part; immoral people will take advantage of the fact that the seller does not have complete information. This could lead to a situation where people who are sicker buy more insurance.

In the case of car insurance, in having full coverage on my piece of crap Audi I would perhaps be inclined to be a tad more reckless (wreck-less?) in going down highways at night. The insurance company would not know that the accident was in fact my own fault. This is the *moral hazard* argument; a party might take unnecessary or silly risks simply because he/she is not going to bear the full costs of a mistake.[1]

Imperfect competition and market failure

(NOTE: It is well worth your while to go through the monopoly diagrams in Theory of the Firm in Chapters 28 and 30.)

As earlier stated; when the market provides the correct amount of a good at the right price, then the basic economic problem has been successfully addressed and resource allocation is optimal. Thus, supply equals demand and marginal cost equals price (or average revenue). As soon as this is not the case, the market fails. This would be the case for firms operating under conditions of *imperfect competition* (monopoly, oligopoly and monopolistic competition) whereby the firm (-s) would be in a position to gain at the expense of the consumer. A firm wielding **monopoly power**, in being threatened by few – if any – viable substitutes on the market, will to a certain extent have the ability to set output and price without consideration of potential rivals. In having this power, the monopolist will certainly have an incentive to lower costs by lowering output, and subsequently set the price as high as possible in order to maximise profits.

This is shown in Figure 21.1. Assume that there are no externalities present and that the monopoly supply and perfectly competitive market (PCM) supply curves would be identical with the MSC curve. The competitive market would produce an output of Q_{PCM} at a price of P_{PCM}. A monopoly, on the other hand, would be prone to increase the price and lower output in order to find the 'greatest distance between price and costs', i.e. maximise profit. When the monopolist raises the price to P_{mon}, the quantity on the market will fall to Q_{mon}. The monopoly has created a market failure since MSB ≠ MSB. The triangle shows that each unit of output between Q_{mon} and Q_{PCM} has a higher marginal social benefit than marginal social cost. (Outside the syllabus: alternatively the net loss of consumer and supplier surplus.) This is the *deadweight loss*.
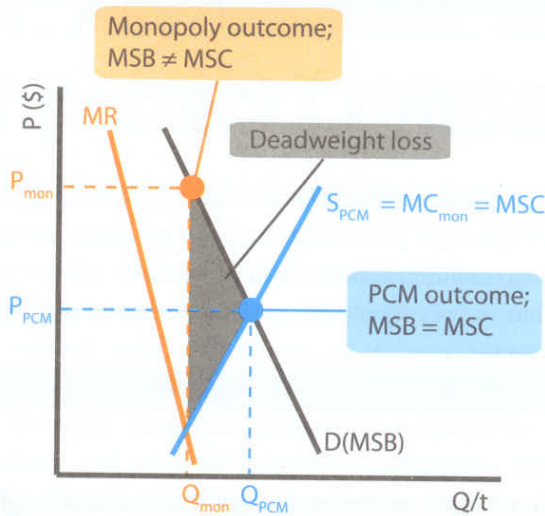
---

1   This was a very weighty argument put forward by economic commentators during the US government's USD85 billion bailout of AIG in 2008. If the US government rescues private financial institutions this would create 'moral hazard' – large financial institutions might increase risky lending practices in the future in the knowledge that government would bail them out if things went wrong.

Monopoly outcome;
MSB ≠ MSC

P ($)

MR

Deadweight loss

P_{mon}

$S_{PCM} = MC_{mon} = MSC$

P_{PCM}

PCM outcome;
MSB = MSC

D(MSB)

Q_{mon} Q_{PCM}

Q/t

Output is set at profitmax level, where MC = MR.
Suboptimal allocative efficiency as P (AR) is not equal to MC.

Misallocation: The net loss of consumer and supplier surplus (grey area) shows the allocative waste: the dead-weight loss.

Figure 21.1 Market failure due to imperfect competition (monopoly)

Imperfect competition is perhaps the most common form of market failure, owing to the strong growth of oligopolies and monopolistic competition in the past 20 or so years. Imperfectly competitive firms – primarily monopolies – can use an assortment of practices which result in market failure. The monopoly can enter new markets by setting low 'predatory prices' which could force other firms out of the market, which would lower competition and give the monopolist more power. The monopoly firm could also set prices at a low level to dissuade any potential entrants from attempting to gain a foothold in the market. Monopolies can restrict market access and competitive markets in many other ways, for example by buying up rivals; owning/controlling vital raw materials; refusing to sell to retail outlets which do not observe a minimum price set by the monopoly; setting different prices on different markets (price discrimination) and finally by simply being able to disregard market demand to a certain degree, since few substitutes are available.

Having said the above, it is quite possible that a monopoly is allocatively *more* efficient than a perfectly competitive market in the presence of high negative externalities. See Chapter 30 for details.

Note that monopolies and near-monopolies can use their market power in many different ways. Firms can attempt to control the market by price fixing (collusion); division of the market in order to limit head-on competition between powerful firms; limiting market entry by withholding access to raw materials or suppliers; forcing suppliers and buyers into disadvantageous contracts; and discriminating against consumers by charging different prices in different locations. All of this serves to enhance the (near-) monopolist's profit at the expense of the consumer and ultimately leads to higher prices and lower output than would be the case in a competitive market.

Summary & revision

1. **Information asymmetry** is a form of market failure that arises when either the seller or buyer in an economic transaction has knowledge/information that the other party does not.

2. **Imperfect competition** leads to market failure since firms in oligopolies, monopolies and monopolistic competition will not produce at the societally optimal level of output. Monopolies are able to set the price above the marginal cost, e.g. MSC > MSB.

1.5

22. Production and Costs

Key concepts: HL extensions

- SR and LR in production
- Total, average and marginal product
- Diminishing returns (revisited)
- Economic costs (opp costs, implicit and explicit components of econ costs)

Fatherly advice to the student!

The theory of the firm is centred on the cost picture of firms and the surrounding environment of competitive, cooperative and interventionist forces. Most of what will be done here will involve a great many diagrams to illustrate and support the narrative. I strongly urge you to do likewise when you address questions and write your Internal Assessment commentaries. The rule in theory of the firm is IF IT CAN BE DONE VIA DIAGRAMS – DO IT!

SR and LR in production

It would be impossible to use absolute time, as in days/weeks/months, to define the short and long run. Any such usage would be rather arbitrary in fact. However, as there is always a time element involved in the speed with which firms can increase output, there is a difference between costs of increasing output. We use this as a device to aid in the definition of short run and long run. For example, it's far easier for a firm to use more electricity and raw material than to increase the size of the production plant – variable factors can and are changed far quicker than fixed factors of production. This is how we can differentiate the short run from the long run:

- The **short run** is defined as the time period where one or more factors of production are fixed, i.e. the period of time within which the firm cannot increase a fixed factor of production.

- As soon as the time period extends to the ability of a firm to increase all factors of production, then the long run has been defined. Quite naturally, in the **long run** all factors and thus all costs are variable.

How long is the long run? Well, how long should a person's legs be?! ('Long enough to reach the ground'[1] – it depends on how tall the person is.) The long run varies immensely depending on the activity of the firm. A nuclear energy firm might have to define the long run as 10 years, which is the time it takes to build a new reactor core and get it up and running. The long run for a pizzeria might be a few months; the time it takes to increase the size of the kitchen and install new ovens. In other words, we don't use the concept of short run as measure of time, but as a definitional period of the length of time needed for a given firm or industry to increase fixed factors of production.

> **Definition: 'Short run and long run'**
>
> The **short run** is defined as the time period during which one or more factors of production are fixed, i.e. there is at least one fixed cost.
>
> The **long run** is when all factors are variable, i.e. all costs are variable.

---

1   Classic answer by Abraham Lincoln.

Total, average and marginal product

Consider the following experiment I've done in class a number of times. Each of my 28 students was instructed on how to fold a specific type of paper airplane which required 12 separate folding movements. The class was then divided in two, 14 students in each group. To make a short story shorter, one group simply made airplanes as individuals, i.e. started and finished each individual airplane by him-/ herself. The other group was asked to divide the labour, where each student would be given one or two folding tasks, pass the paper on to the next student and do the same operation(-s) again with a new paper. Each group was given a stack of paper and 5 minutes of production time. The result was overwhelmingly in favour of the divided labour group. I believe the output was four times higher! A helpful way of rephrasing this is that each individual student produced, on average, four times as many units – which basically meant that productivity was 300% higher. Broadly speaking, the fewer the tasks and actual movements involved per labourer, the higher the productivity is as a whole.

So, if the first student could fold 40 airplanes per hour, adding the second student would mean a total output of 80. The third would bring output to 120, the fourth to 160 and so forth. Right? Nope. This would be to totally disregard the gains to output derived from the division of labour. Two people working and cooperating together create more final output than the sum of two people working separately. Just imagine the time saved – in the paper airplane example – when everyone is fully occupied with a narrow range of simple moves and tasks. No dead or wasted time arises by stooping to pick up another box of paper or leaving the production table to stack airplanes on the shelf. These tasks are already assigned. Output would at first increase at an increasing rate, and then output would increase at a slower rate (diminishing returns sets in) since more and more students are crowding around the same table. The table is a fixed factor of production.

Total and marginal product

What would output look like in a factory where at least one factor of production is fixed? Let us build an example of a small factory making sturdy work-pants. Total product (i.e. total output) is measured per week and we assume labour to be a variable factor while a quantity of capital (here the amount of riveting machines) is fixed. This first labourer could produce 40 Sturdy-Pants within the week…
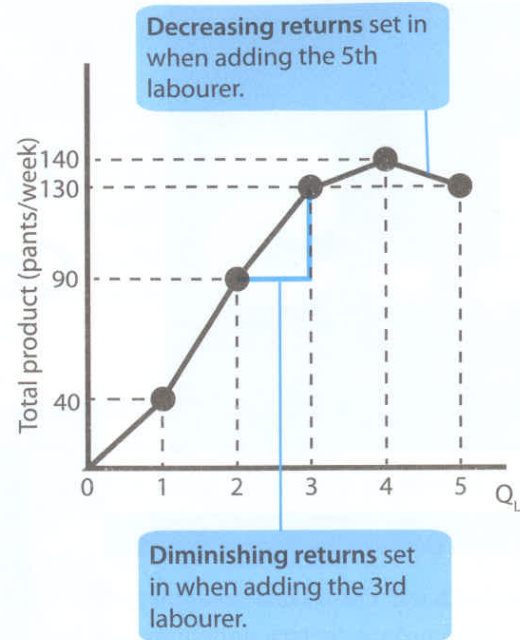
Figure 22.1 Total product (=output) at Sturdy-Pants

…and the second could produce 50, bringing total product (TP) to 90. Figure 22.1 plots out the figures given in Figure 22.2 below, and shows how total product increases by 50 units when adding labourer number 2, i.e. this labourer 'brought with him/her' an additional 50 units. Each additional labourer adds to the total output – this is marginal product (MP). Or, in our example, the correct term would be the marginal output of labour. As tasks are more finely divided and tuned and there is less time-waste, the addition of another unit of labour increases total output more than the previous unit of labour. Each additional labourer 'brings with him/her' more than the previous.

> **Definition: 'Marginal product (output)'**
> The addition to total output caused by adding one more unit of a variable factor (here; labour) is the **marginal product** (output).

Something happens when the 3rd labourer is added. Total product still increases, from 90 pairs of pants per week to 130 per week, but the rate of increase falls – labourer number 2 increased total output by 50 pairs but labourer number 3 increases total product by 40 pairs. It seems that the rate at which total product increases is falling – and labourer number 5 actually decreases total output by 10. (Note: it is always simpler, in concocting tables and diagrams, to put anything 'marginal' [increase or decrease] in between the two total values. I have tried to make this clear by using brackets in Figure 22.3 and then shifting the marginal values down half a step.)

Quantity of labour (Q_L)	Total product (TP)	Average product(AP)		Marginal product (MP)
0	0	-		
				→40
1	40	40		
				→50
2	90	45		
				→40
3	130	43.3		
				→10
4	140	40		
				→−10
5	130	−10		

Average product AP = $\dfrac{TP}{Q_L}$ Marginal product MP = $\dfrac{\Delta TP}{\Delta Q_L}$

Figure 22.2 Total, average and marginal product (output) at Sturdy-Pants

The graph in Figure 22.3 illustrates the 'classic' diminishing return curve. While the first two labourers increase the rate increase of total product, the third labourer lowers – not total product – but the rate of increase in total product. An additional

increase in labour reinforces this - output continues to increase but at an increasingly slower rate. The return on each additional unit of labour is falling; we have diminishing returns. The 4th labourer increases total product by 10 while the 5th labourer actually lowers total output by 10 – in other words, he/she yields negative returns! So too will any additional labourers.

So what is the issue? Do IQs suddenly drop drastically after the first two labourers? No, the answer is that we are operating under the conditions where there are one (or more) factor(s) of production that is fixed. Our premise was that there was a certain amount of *fixed capital*; the riveting machine used to put rivets in stress seams in the pants. As more and more labourers crowd around the machine, at some point they are simply not able to be as efficient in their use of the machine as the previous labourer – thus the machine will not be able to keep up with the relative increase in labour. The machine's capacity remains the same while more and more labourers try to pile around it, each additional labourer beyond number 3 will simply not be able to

Decreasing returns set in when the 5th labourer is added.

The increase in TP – i.e. MP – is maximised when adding the 2nd labourer. This can be seen in the TP curve as the steepest part of the TP slope and in the MP curve as the peak.

Diminishing marginal returns set in when adding the 3rd labourer.

AP will be maximised at the exact point where the AP curve intersects with the MP curve. Thus:

AP_{max} is where MP = AP

This is a very important part of the continuing theoretical rendition.
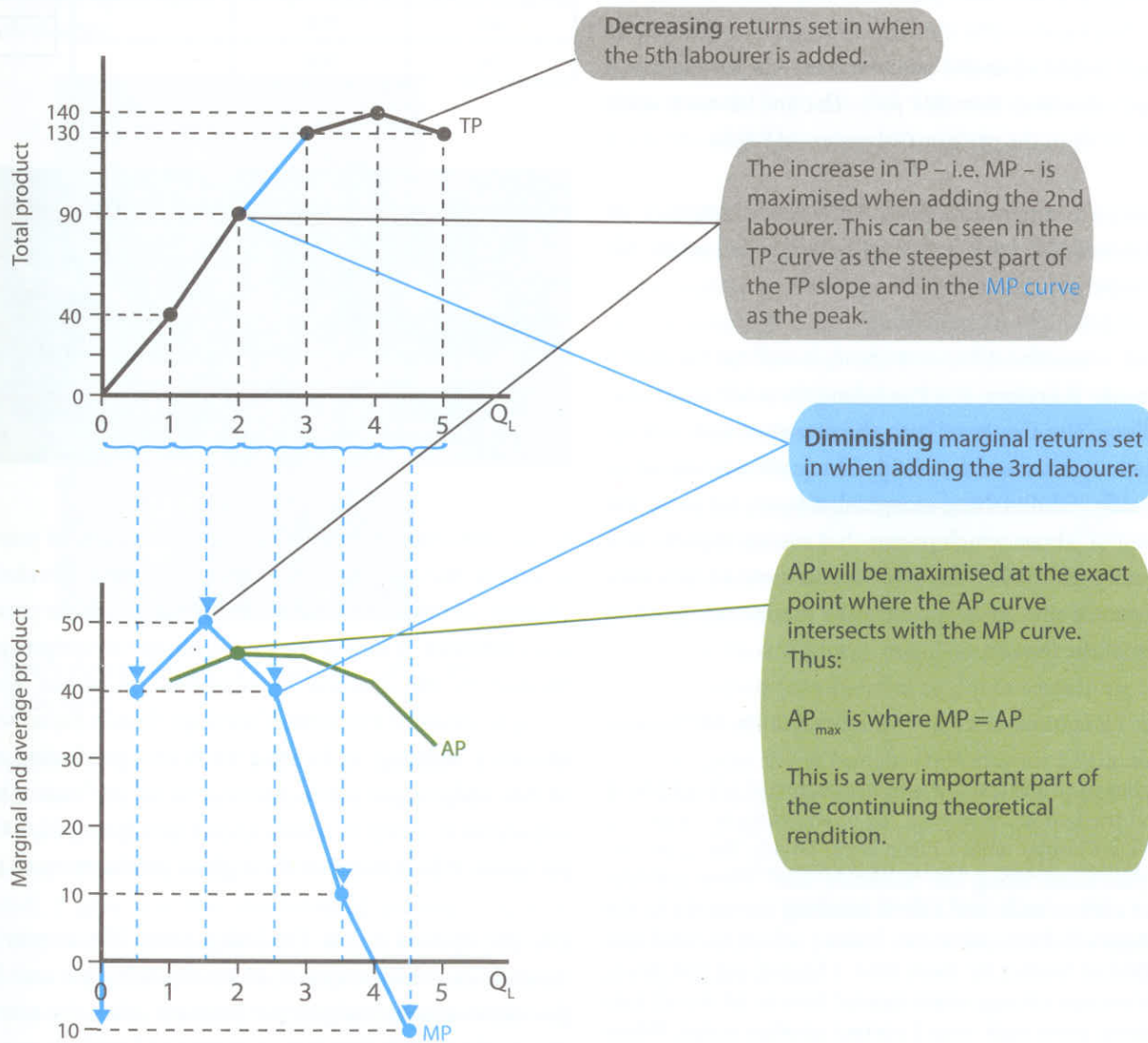
Figure 22.3 Plotting marginal and average product at Sturdy-Pants

utilise the machine to the same extent as previous labourers.[2]

We move on to plotting out marginal and average output in a diagram. The values in Figure 22.2 give the following two curves:

- **Average product:** Total product (TP) divided by the quantity of labour (QL) will render the average product per labourer; this is the grey AP-curve in the lower diagram in Figure 22.3.

- **Marginal product:** The change in TP by adding additional labours is the marginal product (MP); the grey curve in the lower diagram in Figure 22.3. Note that the MP curve is drawn 'in-between' the values for QL in order to show 'change in' or 'addition to' quantity of labour.

There are a few important conclusions to be seen in Figure 22.3 above.

1. The first is that when the rate of increase in TP ('speed of change') *decreases* then *MP falls*. The 3rd labourer adds less to TP than the previous labourer; MP falls.

2. The second conclusion deals with the divergence of the AP and MP curves. MP will start to fall while AP continues to rise, and…

3. …third, when the *AP is at its peak, it will cut the falling MP curve*. Therefore, if AP is falling then MP must also be falling. The shapes of both the marginal and average product curves show that diminishing returns will set in eventually. Diminishing marginal returns set in at the 3rd unit of labour which means that average returns will ultimately also fall. I beg you to take note of this very important point, as it is central to the models used in theory of the firm.

4. When TP is maximised at four labour units, MP is zero.

---

2   To my colleagues: any teacher who has corrected a huge pile of exams and then cooperated with the other teachers correcting the same tests knows what I mean here! Divide the questions up and pass them along the 'assembly line'. Three teachers doing one section each and a forth entering the marks in the grade system will do the job in two hours…where it would take each individual teacher far more time. I figured out that this is primarily because a 8 page exam means I have to set myself into the questions anew each time I correct another script. When there are only 2 pages I can easily recall what the questions ask for and don't have to continuously re-read.

This makes intuitive sense; any additional unit of labour after four will lower TP and of course 'bring with it' a negative value in terms of additional output, e.g. negative MP.

You may be required to do some basic calculations of TP, AP and MP. Note that in the following table, you will get an 'average' value when you calculate MP since the variable factor – labour – is not increasing one unit at a time.

Fill in the blanks…

Quantity of labour (Q$_L$)	Total product (TP)	Average product(AP)	Marginal product (MP)
0	0	-	
10	100		
20	300		
30	600		
40	800		
50	900		

A Little Depth

As an important reminder, any measurement of productivity is simply the quantity of output (or yield), divided by the quantity of input (or production factors). Labour productivity is no different. It means the total quantity of widgets produced divided by the amount of labourers or labour hours. For example, imagine two firms. One large firm in Japan with 2,000 labourers working 40 hours a week and producing a total of 10,000 widgets per week. Another firm, in France, has 4,000 employees working 35 hours a week producing 20,000 widgets per week. Which firm has the highest productivity of labour?!

OK, you spotted it. Yes, I'm being nasty. The answer is that it depends on which measure of productivity we use! If we use the measurement 'widgets per labourer', then the score is one-one, as the French firm has an output of 5 widgets per labourer $\left(\frac{20,000}{5,000}\right)$ while the Japanese firm also has 5 widgets per labourer

$\left(\frac{10,000}{2,000}\right)$. However, if we add in the additional variable of actual amount of time used in production, the labour hours utilised, the score becomes 2 – 1 to the French. The French firm produces the 20,000 widgets using 140,000 labour hours (35 × 4,000), and thus produces widgets at a rate of 7 labour hours per unit $\left(\frac{140,000}{20,000}\right)$. The equivalent rate of labour productivity for Japan is 8 hours per widget $\left(\frac{2,000 \times 40}{10,000}\right)$.

Diminishing returns (revisited)

When I was feverishly trying to get my final paper together for my Master's Degree, my supervisor and I stole a credo (probably from a bumper sticker) to put on the wall of the economics and business institution: 'The first 90% of the job takes 90% of the time. The remaining 10% of the job also takes 90% of the time'. I thought this very funny until it turned out to be true. It became funny again when I realised that it irritated a good many math teacher.

"I've been doing this since I was 20....why can't I do a thousand?"

In adding one more unit of any given variable factor of production, the rate at which output increases will start to decrease – assuming that there is one or more *fixed factors* of production that the firm cannot increase in the given time period. I give you the Diminishing returns Doctrine: 'Diminishing returns ultimately set in if one, or more, factors of production is/are fixed'. Now, how did we previously use '... if one or more factors are fixed'? Think about this while you read on.

Economic costs

I once spent a weekend chopping down oak trees on my grandfather's land in Missouri to sell as firewood. Two days of backbreaking, calloused-hand-inducing hard labour. I sold four cords (a silly measurement used for firewood; about 1.5 × 0.5 × 3 metres) of wood for $140. After subtracting costs of petrol for the pickup truck and chainsaw, two new axe-handles, leather gloves, and a respectable quantity of bandages and iodine (well, I couldn't use the first aid kit at home and let my grandmother know I cut through my foot with the chainsaw) I was left with $110. My costs were thus $30, right?

Wrong. The concept of cost has been used a great deal and now comes the time to more carefully define the term, which unfortunately is from the 'One-plus-one-equals-three' school of mathematics. Costs, according to economists, are simply not a straightforward addition of numbers – that is called accounting cost. Instead, costs in economics take into account the additional *opportunity costs* involved – we *impute* (= assign) *values* to costs which in fact have had no money outlay. Spending the weekend chopping wood meant not being able to work at a local Pimple-Palace (= hamburger restaurant) where I had been offered a job; I was forgoing 18 hours at $3.50 an hour. This $63 in foregone income must be factored-in to the equation of an economist. In addition to this, I could have put the $30 I spent on factor inputs (axe-handles etc.) into something that gave me a rate of return – shares in Apple Computing for example. Let's say that Apple shares bought on Friday for $30 could have been sold on Monday for $31. Finally, there is a risk involved that I would not be able to sell the wood. That risk has a price tag too, which is the minimum return I would require to be in the wood-chopping business in the first place – this is my normal profit, basically the profit I would be satisfied with. Let's impute normal profit at $40.

Here's a summary of the costs:

Accounting costs	
Petrol:	$9
Axe-handles:	$13
Gloves:	$4
First aid:	$4
Sum accounting costs:	**$30**

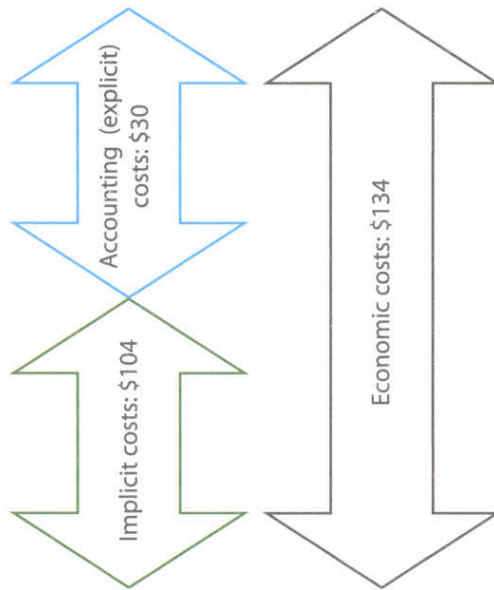| Implicit costs	
Wages:	$63
Share earnings:	$1
Imputed minimum return required (normal profit):	$40
Sum implicit costs:	**$104**
Economic costs:	**$134**

Figure 22.4 Accounting and economic costs for McGee's Chop 'til you Drop

In the example above, we have both explicit (open, clear) and implicit (implied, hidden) opportunity costs:

1. The explicit opportunity costs are of course the accounting costs of petrol, axe-handles, gloves and first aid. The money value of these resources could have been put to other uses.

2. The implicit costs are the opportunity costs of foregone wages, share earnings, and minimum required return on investment.

By adding on the foregone opportunities, which are implicit (= implied) in all economic activity, we get entirely different costs. The implicit costs together with the accounting costs give us **economic costs**…or total opportunity costs of production. I didn't make a profit of $140 minus $30, i.e. $110. According to the reasoning of an economist, I made an economic profit of $140 minus $134, i.e. $6. This silly example is really no different from the way all costs are accounted for by economists. All firms will have alternative uses for land, labour and capital just as in the above example. By using economic costs rather than the more straightforward accounting costs, we are able to assess whether the overall use of resources is optimal or sub-optimal.

Another way of looking at implicit costs is that they are in fact a firm's **normal profit** – the profit necessary to keep the firm in the business. Since I more than covered the implicit costs, I am making an **abnormal** (or **supernormal**) **profit**. We will return to the concept of economic cost and economic profit and expound upon them in the chapter on profit. Just keep in mind that all costs used in economics, fixed or variable, average or marginal, implicitly have economic costs baked into them.

> **Definition: 'Accounting costs and economic costs'**
>
> Accounting costs are visible and/or directly quantifiable costs such as the cost of raw material and labour. Economic costs are estimated by adding onto accounting costs the (imputed – e.g. costs where no payment is made) implicit opportunity costs of using factors of production, plus the "risk premium" of required minimum profit called normal profit.
>
> **Economic costs** = total opportunity costs
>
> **Total opportunity costs** = accounting (explicit) costs + implicit opportunity costs
>
> **Hence, economic costs = explicit + implicit costs**

Do you understand the following joke? If not, revise the section above!

Q: What's the difference between a degree in accounting and a degree in economics?

A: Opportunity Cost!

Summary & revision

1. The **short run** is defined as the time period for a firm where one or more factors of production are fixed.

2. In the **long run**, all factors are variable.

3. **Total product** (TP) is the firm's output during a period of time – in our examples the short run.

4. **Marginal product** (MP) is the addition to total product by increasing a variable factor input by one unit. If labour is the variable input, we get:

$$MP = \frac{\Delta TP}{\Delta Q_L}$$

5. **Average product** (AP) is the is total product divided by the quantity of variable input used, here labour;

$$AP = \frac{TP}{Q_L}$$

6. The MP curve will **intersect the AP curve at maximum AP**.

7. **Diminishing returns** means that output per unit of input will ultimately fall – given that one or more factors are fixed. Diminishing returns is therefore a short run concept and not applicable in the long run.

8. **Economic costs** consist of *explicit costs* (accounting costs) and *implicit costs* (imputed opportunity costs). Explicit costs are costs for raw materials and wages while implicit costs are foregone earnings in the firm's next best alternative in its use of factors of production.

9. When a firm covers both explicit and implicit costs (but no more) then the firm is making a **normal profit**. Thus, implicit costs are in effect the same as normal profit.

23. Short Run Costs

Key concepts: HL extensions

- Short and long run costs
- Total cost picture
- Unit cost picture
- Linking total product to the unit cost picture
- Calculating costs

Short and long run costs

Recall that we define the short run as the time period during which at least one factor of production is fixed – in the long run all factors are variable. The short run is this rather different for a nuclear power plant and the local bakery. It is worth noting, again, that fixed factors are commonly capital or land and that raw materials and labour are frequently used examples of variable factors. However, there are exceptions! Highly specialised labour such as a nuclear physicist or a master pastry chef might well be fixed factors in the examples above. Thus, any **fixed factor** will be a **fixed cost** – and the important thing to keep in mind is that in the short run there is at least one fixed cost since one or more factors of production cannot be increased during the period. When we look beyond the short run in Chapter 24, it will become evident that when we are dealing with the time period where any and all factors of production can be changed we are dealing with the long run.

The iteration so far has dealt with output as a function of the use of factors in the short run. Firms' output decisions will be based on translating factor usage into short run costs figures. We commonly use two 'pictures', i.e. diagrams, to illustrate the costs, the total cost picture and the unit cost picture.

Total cost picture in short run (total variable and fixed costs)

A firm's *total cost picture* is given by taking total fixed costs (**TFC**) and adding the total variable costs (**TVC**). This gives us total costs (**TC**).

$$TC = TFC + TVC$$

It is often easier to see the different costs in a table and diagram. Figure 23.1 shows output and total costs for the Sturdy-Pants Company as used earlier. The quantity is based on the same weekly output figures for the firm. Note that for me to be able to continue with this example, we must make two assumptions:

- We will assume that **only labour costs are variable**. Each unit of labour costs $500 per week.
- **All other factors are fixed** – e.g. everything else is part of the $1,000 'fixed cost package'. (This would mean that raw material, transport, electricity…etc would be fixed – which would most certainly not be the case in a real firm!) Let me once again clarify that examples are often used to help you to understand concepts rather than depict reality.

Total fixed costs are $1,000 per week and labourers have a wage of $500 per unit and week while quantity is at the same output rate as in the previous example. Putting these values into the table gives us the firm's total cost picture.

Q of labour (Q$_L$)	Total output (Q)	Total Fixed Costs ($) (TFC)	Total Variable Costs ($) (TVC)	Total Costs (TC) ($) (= TFC + TVC)
0	0	1,000	0	1,000
1	40	1,000	500	1,500
2	90	1,000	1,000	2,000
3	130	1,000	1,500	2,500
4	140	1,000	2,000	3,000
5	130	1,000	2,500	3,500

Guess what we do now? Yes. Another diagram. This time, however, the object is to show costs, not output. The Y-axis becomes the cost axis while the X-axis is the output axis.

TFC are constant, while TVC and TC rise. The distance between the Q-axis and TFC is of course the same distance between TVC and TC. This is shown by the red double-ended arrows.
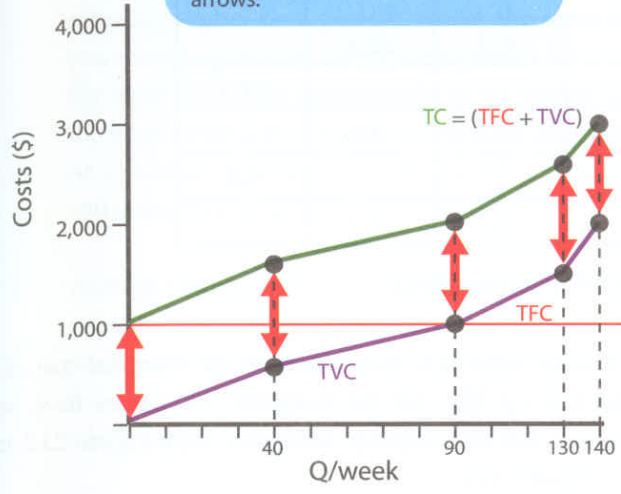

Figure 23.1 Total cost picture at Sturdy-Pants

- **Total fixed costs**: at an output of zero the firms still incurs costs of $1,000. This would be such factor costs as rent, depreciation of capital, insurance and administration personnel – any costs which remain even though the firm is not producing anything.

- **Total variable costs**: as output increases, so does the use of variable factors such as labour (which can also be fixed), raw material, electricity, water, components… etc. (Note that variable costs rise faster after an output of 90 units of pants per week – this is due to *diminishing returns*. See below.)

- **Total costs**: as no surprise, total costs are the sum of total fixed costs and total variable costs – at a quantity of zero total costs are $1,000 (e.g. zero variable costs and only fixed costs) and at 140 units of pants total costs are $3,000 ($1,000 in fixed costs plus $2,000 in variable costs).

Unit cost picture in short run (average and marginal costs)

The total cost curve is just a way of showing the cost of total output at different quantities of pants per week in Figure 23.1 above. Notice that total variable costs (TVC) and thus total costs (TC) increase at an ever slower rate at first but then at an increasing rate. This is in complete correspondence with the law of diminishing returns. We have put a price tag on the use of each unit of labour, and as the marginal output per given unit of labour starts to fall ultimately, then of course the cost of the additional output has to increase. Think of it this way:

- If labourer number 2 is getting $500 and adds 50 pairs of pants to total output, then the marginal cost of each of those pants is $500÷50; **$10** per pair.
- The next labourer adds 40 – but gets the *same wage* – increasing the marginal cost to $500÷40; **$12.5**. Once again, 'diminishing returns' is another way of saying 'increasing marginal costs'.

The above is easier to see in a *unit-cost picture*. Moving from the total cost picture to the unit cost picture is no more of an intellectual jump than moving from total milk consumption in a household to the average quantity per person in the household – and how much milk consumption changes when a friend stays overnight.

Here are the formulae to help you make the jump:

$$TFC + \qquad TVC = \qquad TC$$
$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$
$$\frac{TFC}{Q} = AFC \qquad \frac{TVC}{Q} = AVC \qquad \frac{TC}{Q} = ATC$$

In moving to the unit-cost picture, we divide the totals to arrive at averages.

Marginal values are calculated as:

$$MC = \frac{\Delta TC}{\Delta Q}$$

179

Total output (Q)	TFC	TVC	TC	AFC $\left(\frac{TFC}{Q}\right)$	AVC $\left(\frac{TVC}{Q}\right)$	ATC $\left(\frac{TC}{Q}\right)$	MC $\left(\frac{\Delta TC}{\Delta Q}\right)$
0	$1,000	$0	$1000	-	-	-	
							12.5
40	1,000	500	1,500	$25	$12.5	$37.5	
							10
90	1,000	1,000	2,000	11.1	11.1	22.2	
							12.5
130	1,000	1,500	2,500	7.7	11.5	19.2	
							50
140	1,000	2,000	3,000	7.1	14.3	21.4	

Figure: 23.2 Total and unit costs for Sturdy-Pants

The table above is relatively uncomplicated, yet one point is worthy of comment. That is that it would be most difficult indeed to calculate MC as 'the cost of the last unit produced', which is why calculations are most often based on an average of the last batch, say the final 10 units produced in our example.

The first 40 units have in fact an 'average' marginal cost of $\frac{\$500}{40}$ = $12.5 – e.g. $\frac{\Delta TC}{\Delta Q}$. All the marginal cost values have been calculated like this. Putting the figures from Figure 23.2 into diagrammatic form:
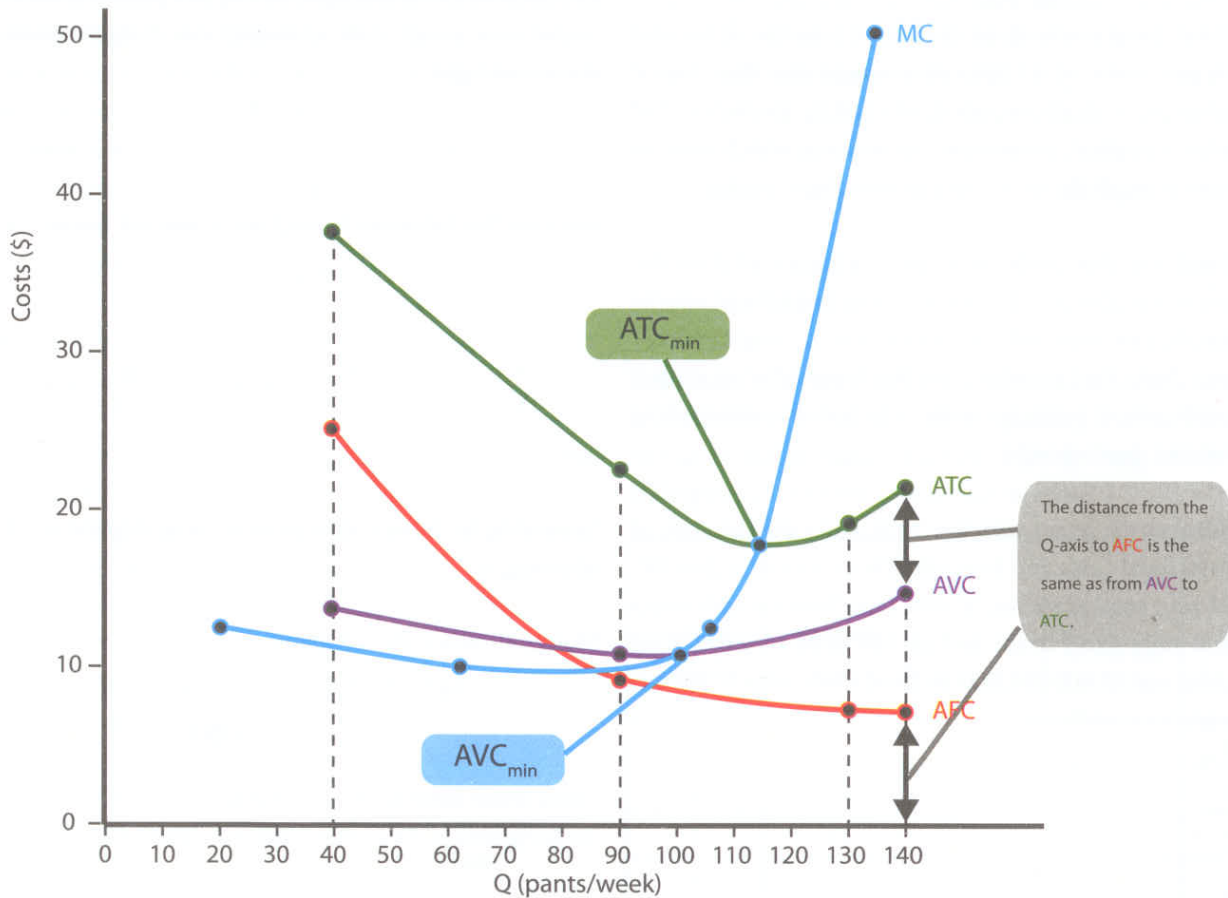

Figure 23.3 The unit cost picture

Note the connection between the MC curve and both the AVC and ATC curves; the MC curve intersects both the average curves' lowest points, i.e. AVC_{min} and ATC_{min}. We will return to this later as it is an important point in our definition of efficiency. Let us now briefly look at each cost curve in Figure 23.4:

- **Average fixed costs:** I often compare TFC with spreading butter on a piece of bread; the more bread you spread a given amount of butter across, the thinner the layer. AFC falls continuously as we are spreading the fixed factor cost of $1,000 across increasing output. At an output of 40 units the AFC is $25 – at output of 140 units, ATC is $7.10.

- **Average variable costs:** The AVC curve falls and then rises as diminishing average returns set in; unit costs rise because increasing the use of variable factors while fixed factors remain constant decreases the rate at which output increases. Average costs rise.

- **Average total costs:** Adding AFC to AVC we get the ATC curve, which also clearly portrays diminishing returns.

- **Marginal costs:** (MC) fall until around 70 units of output and then rise – diminishing marginal returns set in.

Linking total product to the unit cost picture – diminishing returns revisited

We are now back where we started; with diagrams showing how diminishing returns will arise in the short run. The comparison below illustrates how falling marginal output is the same as rising marginal costs by comparing average and marginal product curves with ditto cost curves.

The marginal and average variable cost curves are the other side of the coin, i.e. an inverted version of the marginal and average product curves. In attributing costs to the operation of the firm, the increase and decrease in the rate of production is shown by the U-shaped AC and MC curves.
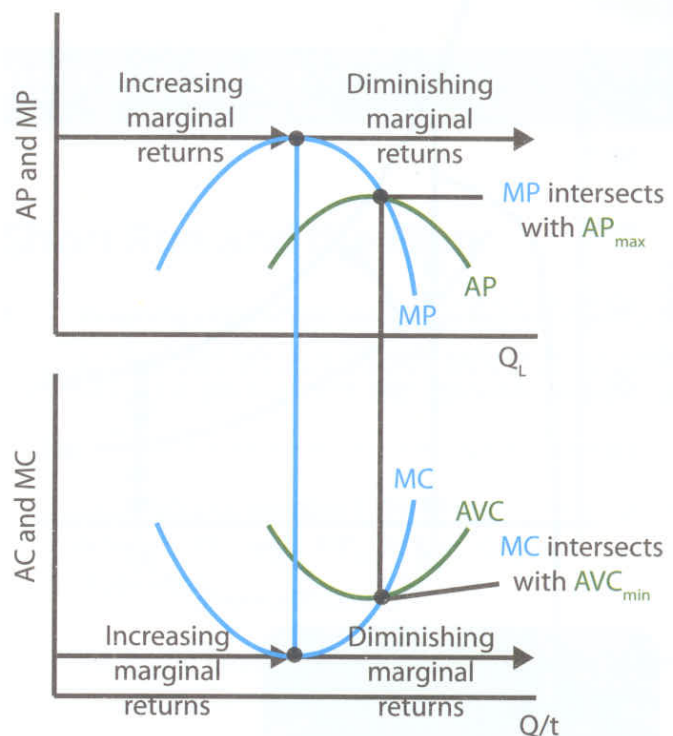
Figure 23.4 Comparing product curves to cost curves

Falling marginal and average output rates and the U-shaped cost curves tell us that diminishing returns are present. Once again; diminishing returns are only possible in the short run – and the short run is defined by the time period during which one or more factors of production are fixed. In the next chapter, we un-fix them….eh, we move into the long run where all factors are variable. Before we get to that, it is necessary to elaborate on the difference between a change in *fixed* costs and a change in *variable* costs.

A change in fixed costs in the short run

Assume that total fixed costs are at first $300,000 and double to £600,000. This means that the average fixed cost per unit goes from $30 at an output of 10,000 units to an AFC of $60. Of course this means that the AFC at 30,000 units goes from $10 to $20.
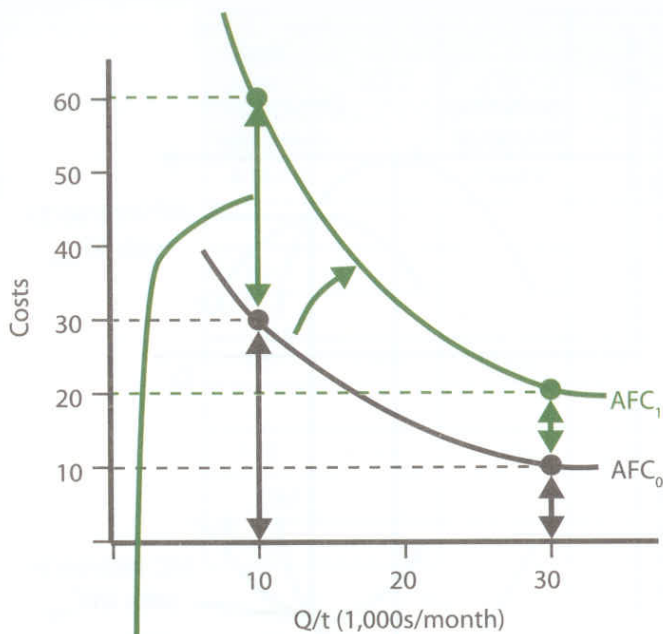
AFC will rise more (in absolute terms) at lower levels of output. This "rolls" the **ATC curve upwards** – as **ATC**$_{min}$ will always intersect with the MC curve.
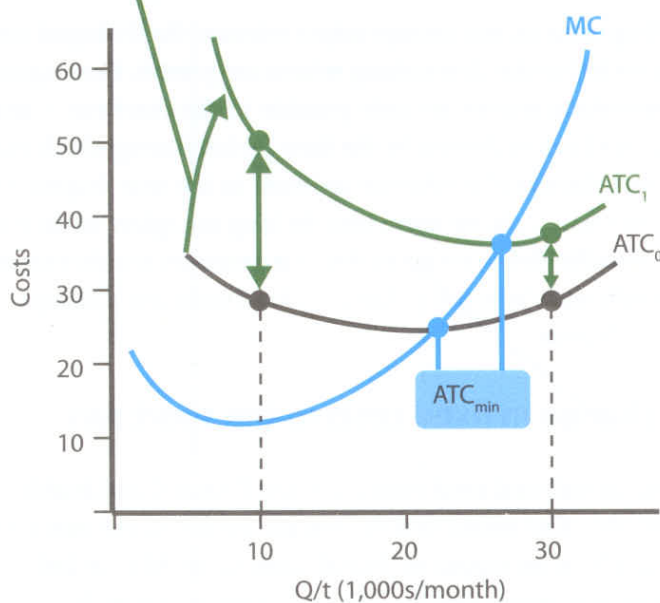
Figure 23.5 An increase in fixed costs – the 'roll' of the ATC curve

Figure 23.5 shows that ATC 'rolls' up the MC curve – the increase in fixed costs causes the AFC curve to shift considerably more at low levels of output than at high ones. An increase in fixed costs – say increased rent – causes the AFC curve to shift upwards. The shift is 'uneven' as fixed costs per unit at low levels will change far more than fixed costs at high levels of output. Therefore, in order for ATC to always intersect with the MC curve when AC is at the lowest point, the ATC curve will 'roll' up the MC curve as shown in the diagram on the left. This is

something many students find confusing when confronted with for the first time; why MC is **unaffected by an increase in fixed costs.**

The reason is that marginal costs measure the change in total costs when an additional unit is produced. The rate of increase remains the same when total costs change by the same amount at all levels. A simple example is given in the table below, where fixed costs double from $100 to $200. This increases total costs at every output level by $100 yet the rate of change – the marginal cost of producing an additional unit – remains the same.[1]

Total output (Q)	TC_0 (FC = $100)	MC_0 ($\Delta TC / \Delta Q$)
0	$100	
		10
1	110	
		8
2	118	
		9
3	127	
		12
4	139	

Total output (Q)	TC_1 (FC = $100)	MC_0 ($\Delta TC / \Delta Q$)
0	$200	
		10
1	210	
		8
2	218	
		9
3	227	
		12
4	239	

A change in variable costs in the short run

A change in variable costs has the effect of shifting the MC curve. If the cost of, say, raw material falls then the cost of producing each additional unit will fall. The result is that the MC curve will shift downwards as in Figure 23.6 below. At every output level the cost of producing an additional unit decreases, shifting the MC curve from MC_0 to MC_1. (Note that in keeping with the axiom that AC_{min} will always intersect with the MC curve, the AC curve will also shift downwards.)

1 Favourite in-class trick: Take three students ranging from small to large, line them up in front of the class and put their heights on the board, e.g. '156 cm → 166 cm → 180 cm'. What's the 'marginal rate of increase in height'? It is 10 cm and 14 cm. "Now guys, stand on your toes...and now stand on a chair... did the marginal values change? Would they change if you were on Mt Everest?" It drives home the point that marginal costs do not change due to a change in fixed costs (disregarding the long run).
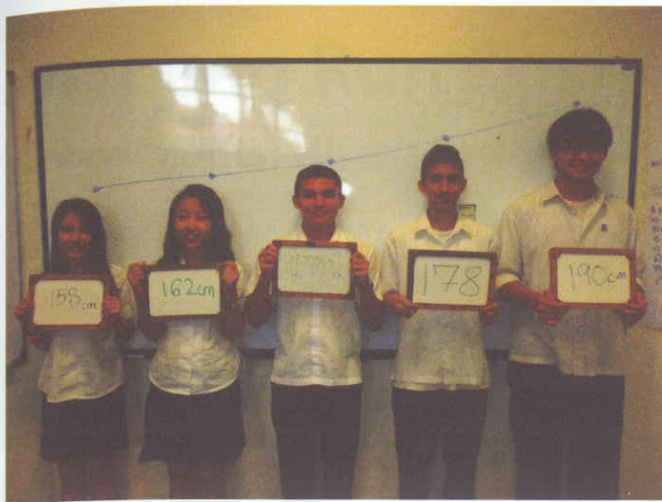
OK guys, now up on your toes...except you Josh
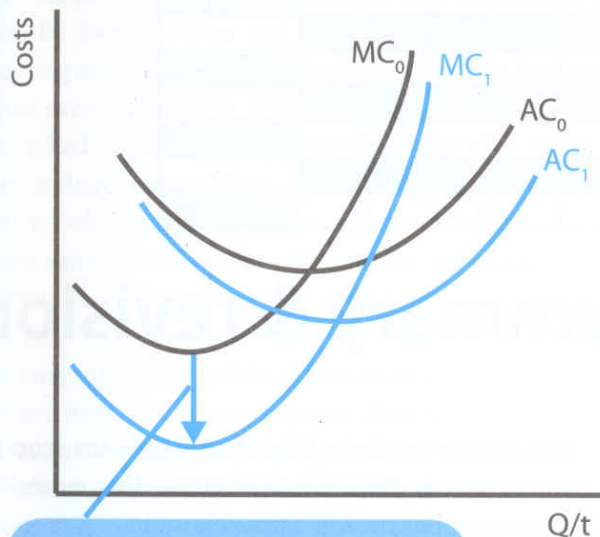
Lower costs of variable inputs (raw material for example) lowers the cost of each additional unit produced → the MC curve shifts down from MC_0 to MC_1.

Figure 23.6 A decrease in variable costs

Question:

What else could cause the MC curve to shift, other than a change in the cost of variable factor inputs?

To summarise, a change in fixed costs will 'roll' the AC curve up or down along the MC curve – always keeping the AC_{min} point at MC – while a change in variable costs will shift the MC curve (and ATC curve). It is worth repeating that we have only dealt with the short run, i.e. one or more factors are fixed.

POP QUIZ 23.1

Short Run and the Firm

1. Assuming a firm where one or more factors of production are fixed and labour is the only variable factor, explain why total product (TP) will ultimately increase at a slower rate.

2. Continuing on question 1 above; what is the connection between TP and MP_{max}? Use diagrams in your explanation.

3. Fill in the blanks: 'When MP turns negative, TP will _____.'

4. Why do marginal and average costs ultimately rise? Would this be equally certain in both the long run and short run?

5. In increasing output from 5,000 units to 5,500 units, TC goes from £100,000 to £130,000. What is the MC per unit of the additional 500 units?

6. Why does a falling marginal product curve show the same thing as a rising marginal cost curve?

7. 'A rise in FC results in a disproportional increase in ATC!' Explain.

9. Explain the effect on MC and AC when FC is tripled. Illustrate with a diagram.

9. Assume that the cost of variable inputs rises for a firm. How will this affect the unit-cost picture of the firm?

Calculating costs

HL students will be required to calculate costs and illustrate them in diagrams in Paper 3. Just bear in mind the following three points:

1. $TC = TFC + TVC$

2. $ATC = AFC + AVC$

3. $MC = \dfrac{\Delta TC}{\Delta Q}$

Total output (Q)	TFC	TVC	TC	AFC	AVC	ATC	MC
0	x	0	x	x	x	x	
							x
10	x	35	x	x	x	x	
							x
20	x	55	x	x	x	x	
							x
30	x	x	x	x	2.667*	x	
							x
40	x	x	x	x	3	x	
							x
50	x	x	200	x	3.8	x	

$\left(\text{*Note; this value is rounded from } 2\tfrac{2}{3}\right).$

The above table shows the total and unit cost picture for a firm. Fill in the missing values (x) and then

a) Draw a total and unit cost picture.

b) Explain why average and marginal costs at first fall and then rise.

c) How would marginal costs be affected if average costs doubles?

Summary & revision

1. Assuming one or more **fixed factors** gives the short run time span, e.g. there are fixed costs. This means that diminishing returns will ultimately set in.

2. The **total cost picture** shows total variable costs, not fixed costs and total costs of production; TVC + TFC = TC.

3. The **unit cost picture** shows average and marginal costs of production – average variable costs, average fixed costs and average total costs. AVC + AFC = ATC.

4. **Marginal cost** (unit cost picture) is the change in total costs over the change in output.

Formulaically; $MC = \dfrac{\Delta TC}{\Delta Q}$.

5. The **marginal cost curve** will always intercept the AVC and ATC curves at their lowest points.

6. **Marginal costs** do not change when fixed costs change.

24. Long Run Costs

Key concepts: HL extensions

- Long run – economies of scale
- Diseconomies of scale
- Long run cost curves

Long run

An American firm was once approached by a Swedish firm proposing a business venture. The American firm politely wrote back asking for references as a guarantee that the Swedish firm had a solid and reputable historical standing and therefore a good credit rating. The Swedish CEO (chief executive officer) promptly wrote back that their credit rating had been excellent during the entire time the firm had been a shareholding company – which started some 40 years before Columbus set off on his voyage to 'India' in 1492. One might say that this is somewhat beyond the boundaries of the short run.

The **long run** is simply defined in economics; it is the time period where a firm can increase any and all factors of production, i.e. **all factors are variable**. Over time, firms will institute changes to factory size, machines, and any other type of fixed factor. Being able to adjust overall factor inputs in proportion to output needs, firms will be able to lower the unit costs as the scale of operations increases.

> **Definition: 'Long run'**
> The **long run** (LR) is defined as the period within which all factors of production are variable. This lowers the average cost since firms will optimise the mixture of fixed and variable factors.

Economies of scale

A firm which increases its use of fixed factors of production will naturally incur higher costs. However, as long as the additional costs are spread over increasing units of output, the unit costs will go down (as long as the proportional increase in unit costs are lower than the increase in output). Say that

a firm is producing 50 thousand units and has total costs of $500 thousand; the average cost per unit is $10. Now, if the firm increases its use of capital, say a new machine in the production line, by a (fixed) cost of $10 thousand and thereby is able to increase output by 5 thousand units, then average cost per unit drops to $9,27 ($510 thousand / 55 thousand units = $9,27 per unit). This fall in long run costs (long run average costs; LRAC) is the result of **economies of scale**, or the **benefits of scale**. Note that 'scale' in the case of firms' operations means size and nothing else.

> **Definition: 'Economies of scale'**
> **Scale economies** are said to exist when a firm increases output and increases both variable and fixed factors, whereby output increases at a proportionately higher rate than the increase in costs per unit. Thus average costs fall as firms achieve increased output per unit of input. This is also known as **benefits of scale.**

As soon as one has grasped the central issue of how increasing costs are spread ever-thinner over increasing quantity of output, it's quite easy to think of real-life examples when scale economies would be apparent; industries which have large fixed costs and/or where there are obvious advantages to being large. Possible scale benefits are often grouped into **internal** and **external** economies of scale.

Internal economies of scale

Internal means that the benefits of being large arise within the firm itself. The increasing scale of operations allows the firm to reap benefits of its own growth – regardless of the overall growth of the industry the firm is in. There are five major sources of internal economies of scale:

1. **Technical economies of scale:** These arise when a firm can take better advantage of its fixed capital. Using machines at 90% of capacity rather than 80% of capacity means that fixed costs per unit produced will fall, since the fixed costs are spread over increasing units. Therefore, buying additional machines and other capital goods serve to lower average costs when the new capital is used to its capacity. This is an example of how **indivisibilities** of production add to the merits of economies of scale; the machine can not be divided in order for the firm to use it on only 4 out of five days (= 80%). The machine will cost the firm just as much in fixed costs on the fifth day, so by utilising the available machine time more fully, the firm will be able to lower costs per unit.

   In addition to production capabilities, increasing size gives firms the ability to fully utilize transport availability. If goods are being shipped by truck, train or cargo ship, the firm will lower transport costs per unit by filling all available shipping space. Another example of decreasing costs of transportation is the **container principle**, which shows that an increase in the surface area of a tanker or other cylindrical container[1] will render a far greater increase in the volume of the container. The increase in available transport volume far exceeds the cost of the increased container size.

2. **Managerial economies of scale:** arise when firms can enhance productivity and output in all areas by increasing specialisation in all areas of production; increasing scale will allow the firm to break tasks down into smaller parts and increase productivity. There are also possible '**synergy**' ('synchronised' plus 'energies') effects; development of new production methods, new materials etc in one area of production can be used across the board in other areas – for example a new plastic manufacturing method in the consumer electronics division could be re-applied to the automotive division.

3. **Purchasing economies of scale:** The larger the firm, the more likely it is to be able to negotiate good contracts for the *bulk-buying* of materials, components and such. Such *purchasing economies* of scale are most noticeable in manufacturing firms which use large amounts of basic commodities. Yet increasingly one finds purchasing economies of scale in firms which actually don't produce any goods themselves; Nike in sports goods and IKEA in furniture are good examples of firms which do not actually produce goods but instead design prototypes and subsequently have manufacturing firms submit lowest-cost tenders for the production of the goods. IKEA basically asks firms to submit 'bids' for the production of a few hundred thousand bookshelves – the lowest bid wins! You can just imagine that IKEA is able to play production companies against each other and get a far lower price per bookshelf than Mom & Pop's Furniture Store up the street.

4. **Marketing economies of scale:** Anyone who's seen TV commercials for large multinational companies in different countries soon notices that the advertisements are often the same no matter where you are. In fact, the commercials are often made without the actors speaking, as this makes it easier to do a voice-over in the background in the local language. Thus, Coke, Honda, Ericsson and Adidas can make one commercial and have it aired all over the world without making 170 different versions. This is an example of *marketing economies* of scale, and also includes such things as having widespread brand recognition, and creating barriers to other firms which do not have the advertising budget of incumbent (= already existing) firms and which are trying to enter the market.

5. **Financial economies of scale:** Getting funds for investment is one of the major issues for firms and the interest payments on loans are a sizable portion of a firm's total costs. Larger firms will often have much more in the way of assets to use as security for loans and therefore can negotiate better terms of interest and repayment than smaller firms. These *financial economies* of scale will give large firms a distinct cost advantage over small firms[2].

---

1 The volume of a container is given by $\pi \times r^2 \times h$. If the radius of a cylindrical container measuring 1 meter times 8 metres increases by 50%, the area of the container increases by 58% but the volume increases by 125%.

2 I often have some fun with the many inventive and often provocative T-shirts worn by my students. I pick someone wearing, say, a red T-shirt emblazoned with the image of Che Guevarra and the old Soviet flag and ask that we go down to the bank (separately) and negotiate a loan. The object is to see who gets the most favourable terms of finance; the Suit and Tie Man (me) or the Provocative T-shirt Person. The issue being dealt with is the perceived risk of the bank in lending money – the lower the lender's perceived risk, the better the terms for the borrower. I believe that lax dress standards helped create the 2007/08 liquidity crisis – you simply should not lend money to a guy wearing flip-flops and a grain store cap.

In addition to the above, there are distinct possibilities of larger firms being able to **spread risks** by diversifying (= broadening) the range of output into other areas. Many Japanese and Korean firms are notable examples of this type of extension. These are often referred to as conglomerates; large firms which produce a vast array of goods – sometimes with no discernable connection between them but often with a good many products which partially overlap in markets. For example, the Korean conglomerate Daewoo International Corporation produces steel, automobiles, media, electronics, textiles, commodities and energy. It has subsidiary companies in over 50 countries and branch offices in most of them. This breadth in terms of both markets and products enables the company to buffer its risks over time and ride out downfalls in regional business activity – the business cycle – and also changes in market demand for products. By having a wide product and sales base, the firm can afford short term losses in one sector by making it up in another.

External economies of scale

When there are benefits to scale arising outside the realm of the firm, one speaks of external benefits of scale. This is the overall growth in an industry which creates benefits by which all firms operating on the market will benefit. Increasing the size of an industry will result in more and better trained labour; more efficient sub-contractors providing parts and components; an infrastructure which benefits all firms in the industry; and a general rise in the knowledge and technology base needed to increase productivity. The Swiss watch industry is a good example, as this very important industry for the Swiss economy has benefited enormously from being concentrated to a limited geographical area, as this has led to a good many external benefits such as a long tradition of schools for watch-making, small parts manufacturing and high quality instruments.

The lure of economies of scale

Scale economies are most valuable in helping to explain a number of events in the global economy during the past 30 to 40 years. Perhaps the most obvious is the increasing concentration of firms, where industries have become increasingly oligopolistic, as firms force out rivals, merge or are bought up by rivals. In the beginning of the 1900s there were over 1,000 automobile manufacturers in the USA, whereas now there are basically three. Mergers and take-overs have been an increasingly strong presence during the 1980s and '90s, and the motives were growth; new markets; attaining new technology and know-how; risk-spreading; and of course to enjoy benefits of scale. This has been most pronounced,

perhaps, in the pharmaceutical business, where soaring costs of producing ever-more complicated and research intensive medicines necessitated consolidation in the industry in order to achieve scale benefits. For example, the merging of Glaxo, SmithKline and Wellcome PLC became what is now the 16th largest company in the world, Glaxo-SmithKline[3].

Going hand-in-glove with 'merger-mania' (as somewhat hyperbolically inclined magazines put it) is the increasing integration of regional markets on a global scale. The rationale is that a small country will have a small market – too small for a firm to get to the level where scale benefits are to be found. The answer is often found in looking to markets outside the country, i.e. exports. Many of the largest multinational companies (MNCs) can be found in very small countries; Holland's Phillips, Finland's Nokia, Switzerland's Nestlé, and Sweden's IKEA to name but a few. It is often argued in today's world of lower barriers to trade and an increasingly integrated global economy that large companies which can receive benefits of scale are the only way to remain competitive.

Diseconomies of scale

Economic theory allows for the possibility of increasing costs in the long run, i.e. where costs per unit increase proportionately more than the increase in output. Such **diseconomies of scale** could result from a number of factors. For example, as organisations and firms become larger, the complexities of management might increase and together with an increasing span of control for middle-management, might result in more time involved in decision-making and implementation. This adds to total costs proportionately more than to output. There might also arise something called *'organisational slack'*, whereby the firm becomes so large that it seemingly lacks real competitors and complacency (= self-satisfaction) sets in, with lower productivity and higher costs as a result.

Please note that I have been most careful to use adjectives such as 'might' and 'could' above. Few studies have shown conclusive evidence for the existence of diseconomies of scale, and I personally file diseconomies of scale under 'Mermaids' or 'Giffen goods'. In reality, firms that might 'outgrow' themselves in terms of factor use versus productivity ultimately scale back on production or cut costs in other ways, thereby avoiding diseconomies of scale.

---

3 *Business Week*, July 14th, 2003; The Global 1000

Definition: 'Diseconomies of scale'

Diseconomies of scale occur when there is a rise in long run average costs – due perhaps to communications problems and lack of control in an increasingly complex organisation.

Long-run cost curves

The central theme in short run costs is productive efficiency, which is defined as how efficiently the firm is using all factors of production – fixed and variable – in producing a given amount of output. This is of course shown by average costs, which fall at first and then rise in accordance with diminishing returns. Optimum productive efficiency exists when the total cost per unit of output is at its lowest. Thus, whereas in the short run AC_{min} would be the measure of efficiency, in the long run we can adjust all factors of production and increase output at a proportionately higher rate than costs. A firm will act in accordance with market demand and estimate whether to increase fixed factors in order to increase output. Figure 24.1 below looks at the possibility for a firm to lower average costs by increasing fixed factors such as machinery or additional production area.

If the firm is producing Q_0 then it would be productively efficient in the short run, since short run average costs (SRAC) are minimised at AC_0. If, however, the firm anticipates that it could indeed sell the quantity Q_1, then it would experience significant increases in average costs as diminishing returns set in, the constraint being the fixed production factors. Therefore the firm assesses options to expand output by increasing certain fixed factors. The estimated average cost of increasing fixed factors is shown by the new short run average cost curve, $SRAC_1$. It is clear that the increased use of fixed factors – new machinery or factory area – adds to average costs at the *present* level of output (Q_0), by 50%, i.e. from AC_0 to AC^*_0. The firm must move to Q_{min} in order to have the same initial cost per unit of AC_0 – and can lower average costs by one third by producing at Q_1, which is the new optimal point of output along the new short-run average cost curve, $SRAC_1$.
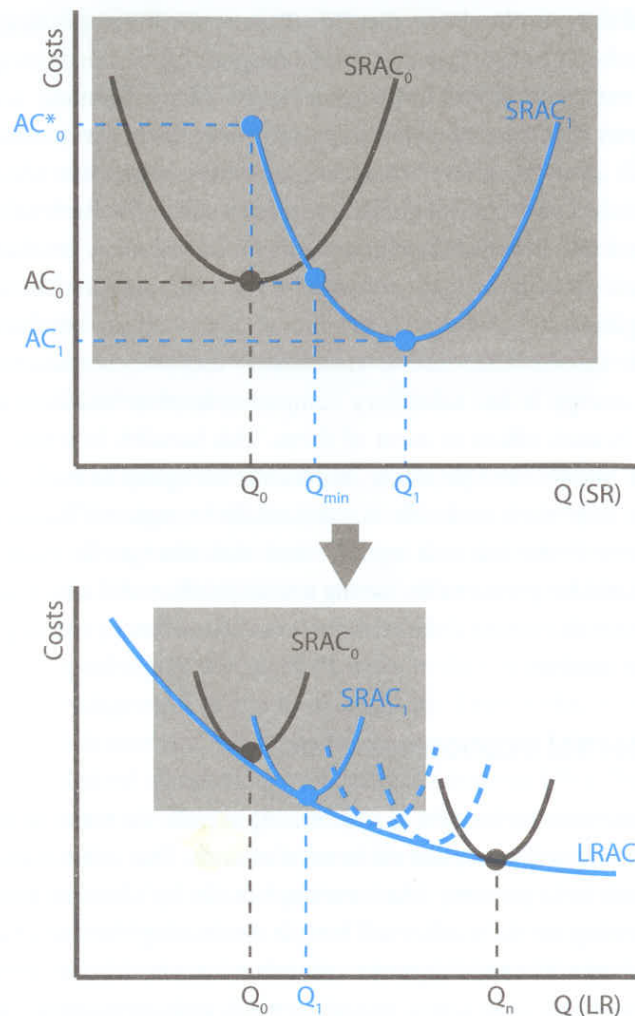
Figure 24.1 Increasing fixed factors – the enveloping effect of average costs in LR

The SRAC curve SRAC0 shows how additional (fixed) factors lift the constraints of the short run, and assuming that any and all additional capital and other fixed factors are variable in the long run, we get an infinite array of SRAC curves, shown in the lower diagram in Figure 24.1. The short run curves are enveloped by the long run curve, or rather, the long run average cost curve is made up of the tangential points of an infinite array of SRAC curves. Note that the assumption of endless SRAC curves means that fixed factors can be increased by minute incremental amounts. Also, any point on the LRAC curve will be equal to or lower than all tangential points on the SRAC curves; the SRAC curves never cross any part of the LRAC as this would mean two possible points on the SRAC curve being equal to the LRAC.

The enveloping effect of the LRAC in Figure 24.2 will form an L-shaped or U-shaped LRAC curve.

- If increased use of fixed factors results in a lower SRAC curve, there are *economies of scale* – or increasing benefits of scale.

- When additional expansion results in same-level SRAC then costs are optimised yielding *constant benefits of scale*. This is the LR optimum scale.
- If SRAC increase this would of mean that average costs are increasing due to the scale of the operation. The firm experiences *diseconomies of scale*. I always show the upward-sloping portion of the LRAC curve as a dotted line to convey the pronounced uncertainty as to the existence of diseconomies of scale. A rather large body of research points to LRAC curves being L-shaped (from Q_0 to Q_1 in the diagram on the left) rather than U-shaped (Q_1 to Q_2) since diseconomies of scale simply are not sufficiently proven.
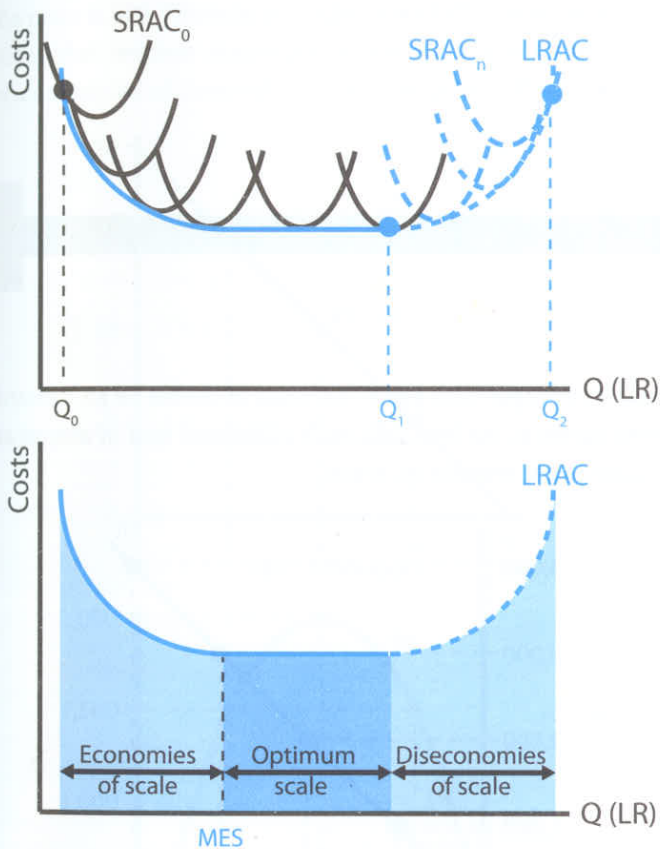

Figure 24.2 LRAC and economies of scale

On a final 'Outside the box' note, Figure 24.2 shows the **minimum efficient scale (MES)** at the point where SRAC curves bottom out. Optimum efficiency in the short run is shown by $SRAC_{min}$, while the long run cost curve has a potential range of **constant benefits of scale**, i.e. many possible points of long run optimum efficiency levels of output. The MES shows the minimum level of output that is possible in order to obtain long run lowest costs in production. We will return to this 'Outside the box' concept in Chapter 85 when we deal with trade barriers in developing countries.

POP QUIZ 24.1

Long Run and the Firm

1. For what reasons might (internal) economies of scale exist?

2. How might the existence of economies of scale be linked to; a) the increasing size of firms; b) mergers; c) oligopolies; d) increasing trade and globalisation?

3. Using diagrams, explain carefully the difference between diminishing returns and diseconomies of scale.

4. Why is AC used as a measure of productive efficiency? Why not MC?!

Summary & revision

1. The **long run** is the time period for a firm during which all factors are variable.

2. When a firm increases both variable and fixed factors and output rises faster that total costs, average costs fall – **economies of scale** exist.

3. When long run costs rise proportionately to output, there are **constant benefits of scale.** While unlikely, it is possible that average costs in the long run rise faster than output in which case the firm has hit **diseconomies of scale.**

4. **Internal economies of scale** arise due to *lower production costs* resulting from new technology, sharing of knowledge, spread of admin costs and bulk-buying of raw materials and components; *marketing* economies of scale such as brand recognition; and *financial* economies of scale when a firm gets easier and/or cheaper credit.

5. **External economies of scale** arise outside the firm and are created when a large industry is created and individual firms enjoy a larger pool of (specialised) labour, better communications and infrastructure and possibly related educational facilities for R&D.

25. Theory of the Firm – Revenues and Profit

Key concepts: HL extensions

- Total revenue
  - o Perfectly competitive market firm
  - o Monopoly firm
- Marginal revenue
  - o Perfectly competitive market firm
  - o Monopoly firm
- Average revenue
  - o Perfectly competitive market firm
  - o Monopoly firm
- Profit and loss
- Calculating revenues and profit

Total revenue

Total revenue (TR) is the 'money coming in' to the firm by way of sold goods – e.g. price times quantity sold. What happens to total revenue as output increases or decreases depends on whether the firm has pricing power or not. Let us compare the perfectly competitive market firm (price-taker) with the monopoly (price setter) with regards to total, marginal and average revenue.

Perfectly competitive market firm

TR for a competitive firm

Assume that the market for Widgets is instead a perfectly competitive market (definition coming up), and we turn our attention to one single firm in the market. Assume that there are 1,000 firms supplying the Widget market and none of them is big enough to influence supply – or price. The firm is thus a *price-taker* and has simply to accept a market price. Now, a little bit of syllogistic logic (see 'syllogism' Chapter 2) says that if the firm is too small to influence supply, and if the firm simply accepts the market price, then the firm can sell every unit it produces at the going market price.

Figure 25.1 shows the total revenue curve for a firm operating under perfectly competitive market conditions. Assuming that the market price is $10, total revenue is a matter of multiplying

quantity of output with price. TR must of course be an *upward sloping* curve; as the firm sells each additional unit of output at the same price, revenue increases.
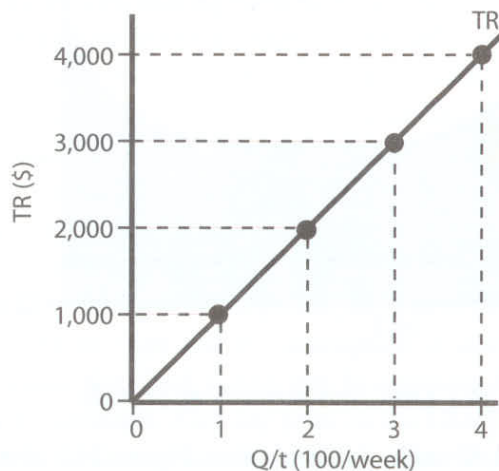
A firm operating under conditions of competitive markets is too small to influence total market output. It can sell any amount it produces and is unable to influence the market price. TR increases when the firm sells more units, giving an upward sloping TR curve.

Figure 25.1 Total revenue (TR) for a firm operating on a perfectly competitive market

Monopoly firm

Note that total revenue shown earlier in the supply and demand model (Chapter 9) was for the *market* as a whole and that the TR curve was defined by the boundary of the demand curve. Assume now that the market demand is supplied in total by a *single firm*, a monopoly.

TR for a monopoly firm

As we are assuming that only one firm supplies the market, market demand will be the same for the monopoly firm. The firm can set any price and the pattern of demand will determine output. In Figure 25.2 total revenue will peak at $2,000, when the price is $10. This is in accordance with the earlier rule which states that revenue maximum is where price is set at unitary price elasticity of demand. At any other price, TR will be lower.
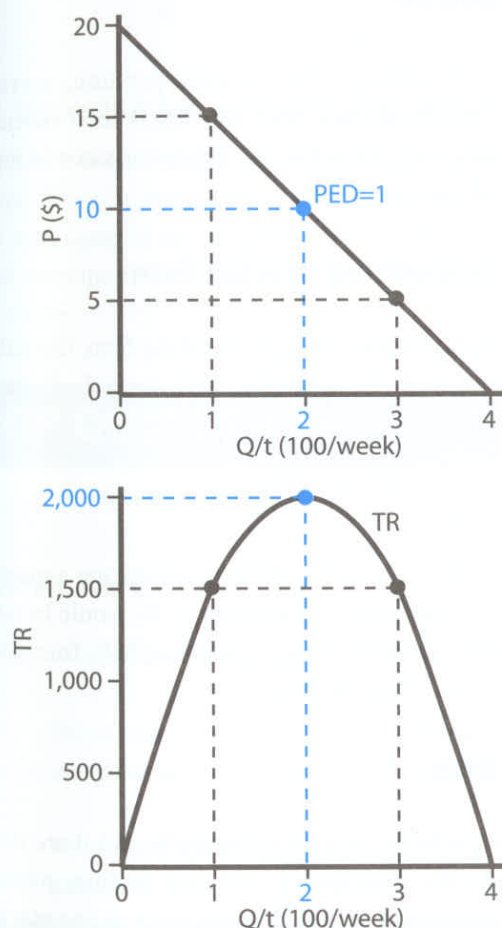
At a price of $10, PED is unitary, which means that total revenue will be maximised at this price.
WARNING!
It is usually around this time that students start to equate revenue with profit. Don't! This is an all too common and grave error. We will look into profit shortly. Chill.

Figure 25.2 Total revenue (TR) for a monopoly

Marginal revenue

Recall that the definition of marginal cost is the change in total cost over the change in output. Marginal revenue is the same; it is the addition to total revenue over the change in quantity sold. We often put this in terms of the revenue gained in selling one more unit, or, but formulaically speaking we get;

$$MR = \frac{\Delta TR}{\Delta Q}$$

Perfectly competitive market firm

Again, the perfectly competitive market firm is a price-*taker* so it will simply accept the market price. Figure 25.3 shows that a price of $10 per unit adds to total revenue by the same amount as the price – selling one more unit at a price of $10 adds the same amount total revenue. The rule is summarised as '**P = MR for a perfectly competitive firm**'.
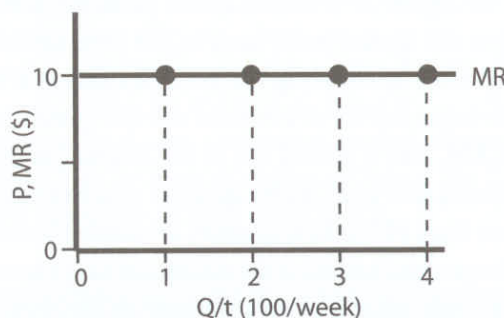
A competitive market firm is a price-taker. A price of $10 means that every additional unit sold adds to TR by the same amount – which is, of course, marginal revenue (MR = ΔTR/ΔQ).

Figure 25.3 Marginal revenue (MR) for a firm operating on a perfectly competitive market

Monopoly firm

As simple as marginal revenue is to grasp, as incomprehensible at first sight is the concept when applied to a monopoly firm. Let us lay down the relevant assumptions again; the monopoly firm is alone in the market and is able to set the market price – therefore the market demand is the same as the monopoly's demand. We also assume that the monopoly sets a single price (the relevance of this assumption will become clear in the chapter on price discrimination). Using the same ballpark figures as for the competitive market firm described above gives Figure 25.4:

P ($)	Q	TR ($)
20	0	0
15	1	15
10	2	20
5	3	15
0	4	0

MR ($)
15
5
−5
−15

Figure 25.4 Marginal revenue table for a monopoly

Two things stand out immediately.

1. Firstly, MR is always lower than the price (which is the same as average revenue)

2. MR is continuously decreasing throughout the span of the demand curve.

This gives us Figure 25.5 below, which plots the MR of the monopoly. As the market is the same as the firm, basically, the firm will sell more by lowering the price. Starting at a price of $20 and moving to a price of $15, total revenue goes from $0 to $ 15; MR = $15. This is plotted out in the diagram as the point halfway between zero and 1 on the quantity axis. In lowering the price further, from $15 to $10, quantity demanded increases by one unit; the revenue gained is the additional $10. However, as the firm is a single-price monopolist (more on this further on), it must sell both the first and second unit for $10, bringing TR up to $20. This is a $5 increase over the previous $15, leading to a MR of $5, shown by the second point between 1 and 2 on the quantity axis.
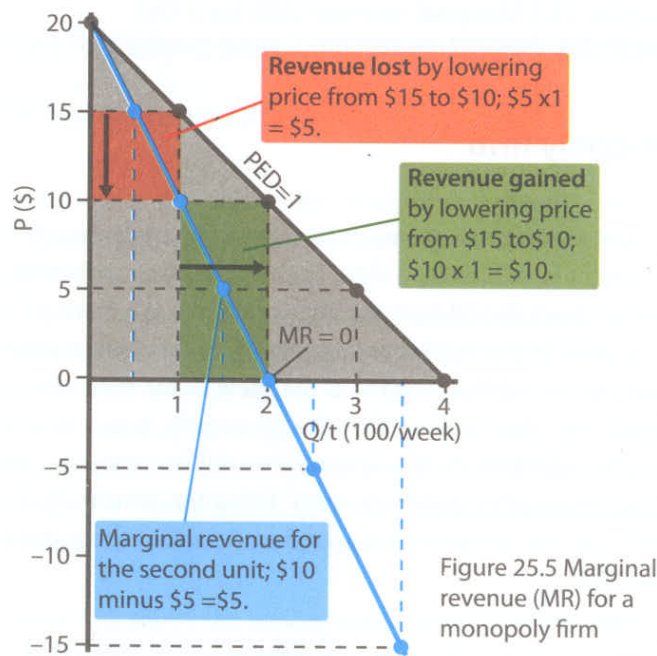
Figure 25.5 Marginal revenue (MR) for a monopoly firm

Perhaps a simpler way of looking at MR for a monopolist is to look at the *areas* of revenue gain and revenue loss. Going, once again, from a price of $15 to $10, the revenue gain is the blue area in Figure 25.5. The subsequent loss of revenue is the light grey area; the difference is MR. Doing the same all along the demand curve renders the MR curve for the monopolist – it is 'halfway' between the demand curve and the price axis. Now, fill in the following blanks below as you wait with baited breath for greater depth in the chapter on monopoly:

I) 'When PED is unitary, the MR is _____!'

II) 'TR is maximised when MR is _____!'

III) 'As TR increases, MR is _____!'

Average revenue

The last version of revenue is the revenue per unit; average revenue, or AR. As in average cost, the AR is total revenue divided by quantity sold; AR = TR / Q. This is the same in both a competitive and monopoly firm.

Perfectly competitive market firm

Just as MR was equal to price for a price-taking firm, the AR is equal to price. This can be shown by breaking down the formula for AR.

$$AR = \frac{TR}{Q} \rightarrow \frac{P \times Q}{Q} \rightarrow \frac{P \times \cancel{Q}}{\cancel{Q}} \rightarrow AR = P$$

In the example used earlier, when the firm was facing a market price of $10 and sold a quantity of 4 Widgets, TR would be $40. Dividing this by the quantity of 4 gives an AR of $10. Therefore, in a price-taking firm, P = MR = AR.

Monopoly firm

The above holds perfectly true for a monopoly, yet there is an additional link to the monopoly's unit cost picture, namely that the average revenue curve will be the same as the market demand curve. Figure 25.6 below shows that even though TR rises and then falls, AR will follow the same path as the demand curve. The conclusion is that **in a monopoly firm, P = AR = D.**

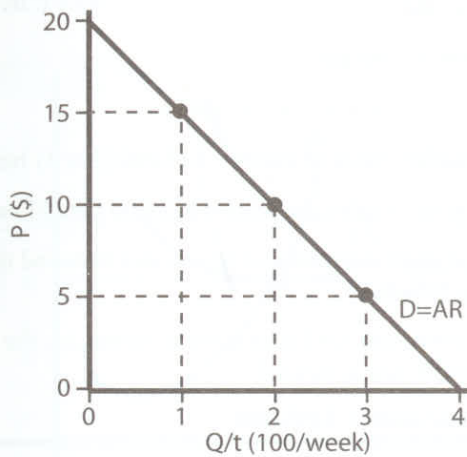| P	Q	TR ($)	AR ($)
20	0	0	-
15	1	15	15
10	2	20	10
5	3	15	5
0	4	0	0

Figure 25.6 Average revenue (AR) for a monopoly

By now it is quite easy to have forgotten or confused everything done above in terms of revenues! Don't worry, we're going to do it all again in the chapters on perfectly competitive markets and monopolies.

POP QUIZ 25.1

Revenue

1. When does the maxim 'P = MR' hold true?

2. Show, using diagrams, the difference between the marginal revenue of a competitive firm and a monopoly.

3. What happens to TR in a monopoly as MR values turn negative?

Profit and loss

At a basic level, profit is total revenue minus total costs. However, it is most important, as an economist, that you understand the monumental difference between accounting profit done in business studies and the concept of **economic profit**. We have dealt with the concept of economic costs using the basic concept of opportunity cost and now we take it one step further by adding on revenue and profit.

Basically, **normal profit** is when a business or enterprise uses resources and receives an amount of revenue or income that gives him/her a rate of return that is no different from the next best alternative. If the venture gives a rate of return that is higher, we speak of **economic/abnormal/ supernormal profits**.

> **Definition: 'Normal and supernormal profit'**
> **Normal profit** is the profit earned by the firm which covers all costs – both accounting and opportunity costs – and hence TC = TR. Thus there is no incentive for the firm to leave the market as the second best alternative cannot increase profits. This profit enables/incentivises the firm to keep resources in their present use.
>
> Any profit above and beyond normal profit is **supernormal profit** (or economic, abnormal, excess profit) – this creates incentives for market entry by other firms.

Let's say I finally manage to get myself severely, unalterably and definitively fired, and then pursue my passion for Swiss watches by starting a fine watch shop. Below I give you the accountant's version and the economist's version of the same figures. (I have colour coded these costs in terms of *implicit* and *explicit* costs – see footnote for key.[1])

---

1 Explicit costs: red. Implicit costs: green.

The accounting method	The economist's method
TR = £500 000	TR = £500 000
Costs:	Costs:
- Wholesale cost of watches = £300,000	- Wholesale cost of watches = £300,000
- Utilities, services, insurance = £40,000	- Utilities, services, insurance = £40 000
- Wages = £40,000	- Wages = £40,000
- Depreciation = £30,000	- Depreciation = £30,000
- Interest paid on bank loans = £20,000	- Interest paid on bank loans = £20,000
	1) - Fall in market value of assets = £15,000 [See 1) below]
	2) - My foregone wages (implicit) = £25,000 [See 2) below]
	3) - Foregone interest revenue of my money invested in firm = £15,000 [See 3) below]
	4) - My 'risk premium' for being in the business, i.e. the normal profit = £10,000 [See 4) below]
Total costs = £430,000	Total economic costs = £495,000
Profit = £70,000	Supernormal profit = £5,000
	(or economic/abnormal/excess profit)

Comments on the additional costs in the economist's version:

1. **Cost of keeping a stock of watches:** There is an estimated opportunity cost of £15 000 in keeping a supply of watches in stock as that money could have been used otherwise.

2. **Foregone wages (implicit):** I could have made £25 000 in my next best option available and therefore there is an imputed opportunity cost of that income foregone.

3. **Foregone interest revenue of my own money invested in firm:** I invested £300 000 in the firm and current interest would yield 5% per annum; 0,05 x £300 000 = £15 000.

4. My 'risk premium': There is always a risk of losing everything by being in business for oneself. The amount that I estimate the risk at is £10 000, which is to say that my estimate of the lower limit of profit in undertaking the watch enterprise is £10 000 – my normal profit. Simply put, I estimate that the opportunity costs of taking the risk at £10,000.

What if...

1. ...my foregone wages were £30,000? I would have total revenue equalling total costs (both explicit and implicit) and thus 'only' a **normal profit**. Note the quotation marks around the term 'only'! It is a common misconception to regard normal profit as 'not too good' or 'substandard'. Not true; a normal profit means the firm is covering all its opportunity costs and simply cannot do any better – as measured by the next best opportunity foregone.

2. ...my foregone wages were £31,000? Total revenue would be exceeded by total costs by £1,000 and I would be incurring a **loss**. I would do better by taking my resources and putting them into the next best alternative – e.g. I would leave the market.

In summary: whenever we in the future talk about 'profit', we are inferring that both the opportunity cost issues and imputed risk premium – normal profits – have in some way been implicitly baked into the cost picture of the firm. 'Profit' in economics is anything above and beyond covering opportunity costs and risk premiums, i.e. 'abnormal /supernormal profit' unless otherwise specified.

Calculating revenues and profit

HL students will be required to calculate both costs and revenue in order to calculate profit and illustrate them in diagrams in Paper 3. At the risk of beating you over the head with it again; *profitmax* output is set at MC = MR and *total profit* is calculated as (AR − AC) × Q.
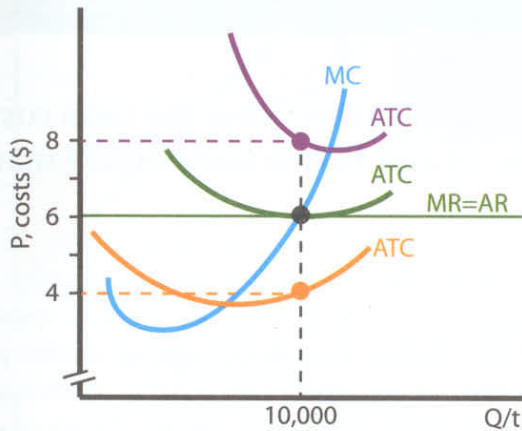

Figure 25.6

Figure 25.6 above shows the unit cost picture for a competitive market firm. For each AC curve, calculate profit/loss. Let's check your insight here: why doesn't the firm change output for each different AC curve?

Summary & revision

1. **Total revenue** is defined as price times quantity;
    TR = P × Q

    a. For a *competitive* market firm (price taker!), the TR curve will be linear, upward sloping and starting at origin

    b. For a *monopoly* firm (price setter!), total revenue along a downward sloping demand curve increases to where PED is unitary and then falls.

2. **Marginal revenue** is the change in total revenue over the change in quantity;

    $$MR = \frac{\Delta TR}{\Delta Q}$$

    a. For a *competitive* market firm, the marginal revenue will be the same as price, e.g. an increase of one unit sold will increase total revenue by the same as price.

    b. For a *monopoly* firm, marginal revenue will fall faster than average revenue – the MR curve will be twice as steep as the AR (D) curve.

3. **Average revenue** is total revenue divided by quantity of output;

    $$AR = \frac{TR}{Q}$$

    a. In a *competitive* market firm, average revenue will be the same as the price and marginal revenue;

    $$AR = \frac{TR}{Q} \rightarrow \frac{P \times Q}{Q} \rightarrow \frac{P \times Q}{Q} \rightarrow AR = P$$

    Thus: for a competitive market firm, P = MR = AR

    b. In a *monopoly*, AR is also the same as the price but in being a price setter the monopoly sets price or output along its own demand curve. Thus, AR is the same as the demand curve.

4. **Profit** in economics is referred to as '**surpernormal**' or '**abnormal**' profit. This is when total revenue exceeds total costs. Total costs include the explicit (accounting) costs and the implicit costs of production – these are the *economic* costs.

5. **Normal profit** is when total costs (including all implicit costs!) equal total revenue.

26. Goals of Firms

Key concepts: HL extensions

- Profit maximisation
- Alternative goals of firms

Profit maximisation

Every year when I get to this point in the syllabus, I have threatened to have a set of T-shirts made for my people with the legend 'MC = MR Rules OK!' emblazoned on the front – and maybe 'MR MCgee Rules Too!' on the back. That would wake the local school board up in the morning. What does the legend mean? It means that profit is maximised at the point where marginal cost is equal to marginal revenue (but only when MCs are rising, see following). This is the short run **profitmax condition** facing firms in the short run.

> **Definition: 'MC = MR profitmax condition'**
> Profit will be maximised when the cost of the last unit equals the revenue of the last unit. This point of output is where MC = MR.

Profit maximisation using the total cost picture; TR and TC in a competitive market firm

Let's use a scale which is a bit more realistic. Assume a Widget firm (Widget = 'manufactured thing') which is operating in a competitive market and cannot influence market price or market output – it will produce the quantity which maximises profit as we assume that the firm is a short run profit maximiser. The total cost picture and equivalent unit-cost picture are given in Figure 26.1. The total cost picture shows increasing TR as more units are sold at the market price of $5, while the TC curve portrays the standard diminishing return curvature. At zero output, the firm would make a loss of $5,000, as the fixed costs would not be covered by any revenue. As output increases and revenue does too, total costs increase at a slower rate giving the firm a break-even point (where TR = TC) at 1,500 units. This would give the firm a **normal profit**.

The area shaded in blue shows the supernormal profit (remember, opportunity costs are included in TC!) attainable at output levels between 1,500 and 4,000 Widgets per day. Supernormal profit is maximised – i.e. the difference between TR and TC is the greatest – when the firm is producing 3,000 units a day. At this output level the TR is $15,000 and TC is $10,000, yielding a supernormal profit of $5,000 per day. It is worthy of notice that should the firm decide to forgo profit maximisation, perhaps in order to expand its sphere to new markets, it could produce an additional 1,000 units and set output level at 4,000 units where, once again, TC is equal to TR.
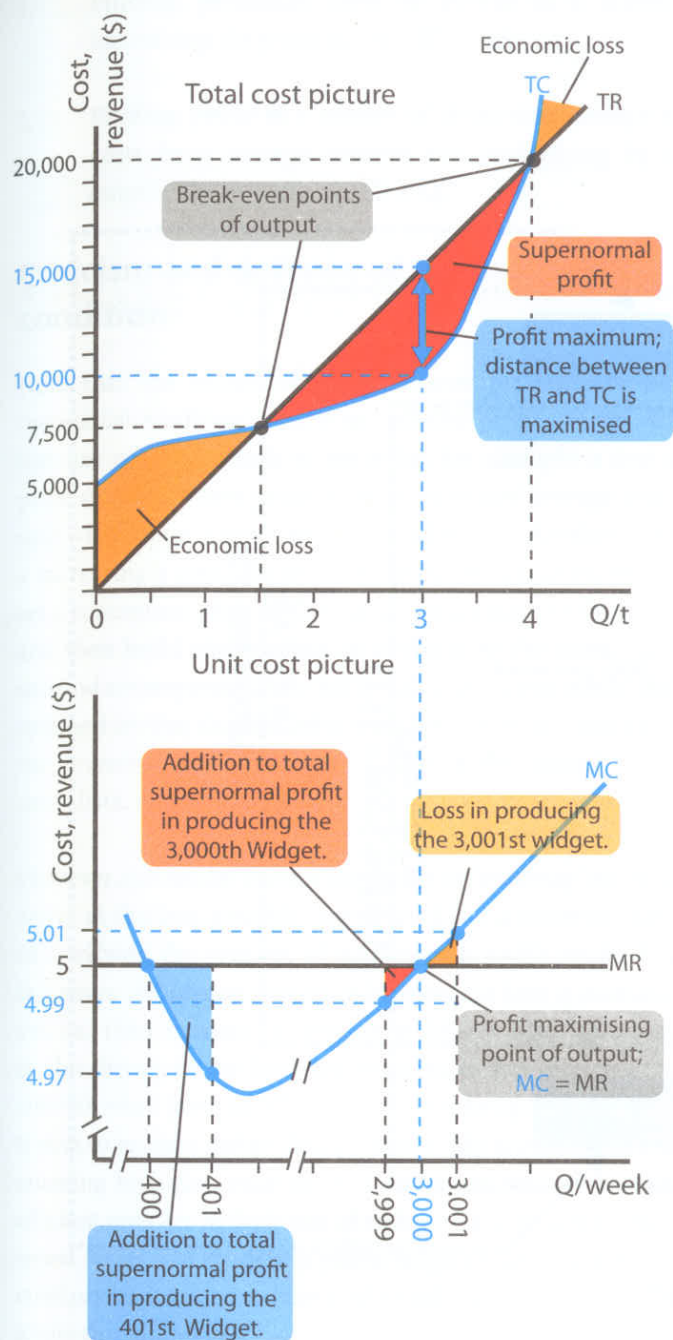
The curves in the unit cost picture show that the first 399 units will cost more to produce than the firm can sell them for at the market price of $5. Producing the 401st unit would add to total costs by $4.97 and add $5 towards total revenue, giving $0.03 towards profit. Note that this is NOT profit but a *contribution* to making up for the initial loss making production of the first units. Subsequent units will have lower marginal costs (MC) than the market price of $5; i.e. each unit will have a MR of $5 and even after diminishing marginal returns kick in and the contributions to total profit become ever-smaller, the firm still stands to gain by producing them. This is why the 3,000th Widget is produced; it adds one cent in contributing to total profits. (Note that I have broken the P-axis in several places.) Producing the 3,001st Widget would cost $5.01 but still only fetch the market price of $5, which would remove 1 cent from overall profit.

We now have the basis for the profitmax condition facing a short run profit maximising firm. A firm will expand output beyond initial loss-making levels as marginal costs fall to below market price. The profit-maximising firm will **increase output to the quantity where MC = MR**[1] but **decrease output if MC > MR**. This sets the profit maximising level of output at the point where marginal costs are equal to marginal revenue.

Definition: 'Profitmax condition of MC = MR'

When MC < MR a profit maximising firm will increase output. If MC > MR (and MC is rising), then output will be decreased. The profit maximising level of output is when MC = MR (and MC is rising).

Profit in the unit cost picture; (AR – ATC) x Q

Now, using the unit cost picture in Figure 26.1 above, calculate profit. No, don't bother, I'm being silly – and wasting your time! You cannot calculate or show profit without an ATC curve of course. Recall that profit is total revenue minus total costs. Therefore, in our unit cost picture, profit is average revenue minus average total costs (times quantity).

Figure 26.1 Profit maximisation using TR/TC and MR/MC curves

Economic theory predicts (as does mathematical reasoning, as soon as the unit cost concepts are understood) that if the price one can get for the sale of an additional unit (MR) is higher than the cost of producing that unit (MC), then it will be produced. This is a somewhat simplistic way of introducing the MC = MR maxim (= saying, proverb), but it is a common sense argument readily understood. The unit-cost diagram in Figure 26.1 shows an 'abbreviated' (and grossly exaggerated!) MC curve in order to be able to fit the profitmax level of output in conjunction with the total-cost picture above. (Note that the Q-axis is not equidistant to the total-cost picture.)

---

1   Note that there are in fact *two points where MC = MR*, so one should add the qualification that profitmax output is where MC = MR when *marginal costs are rising*.
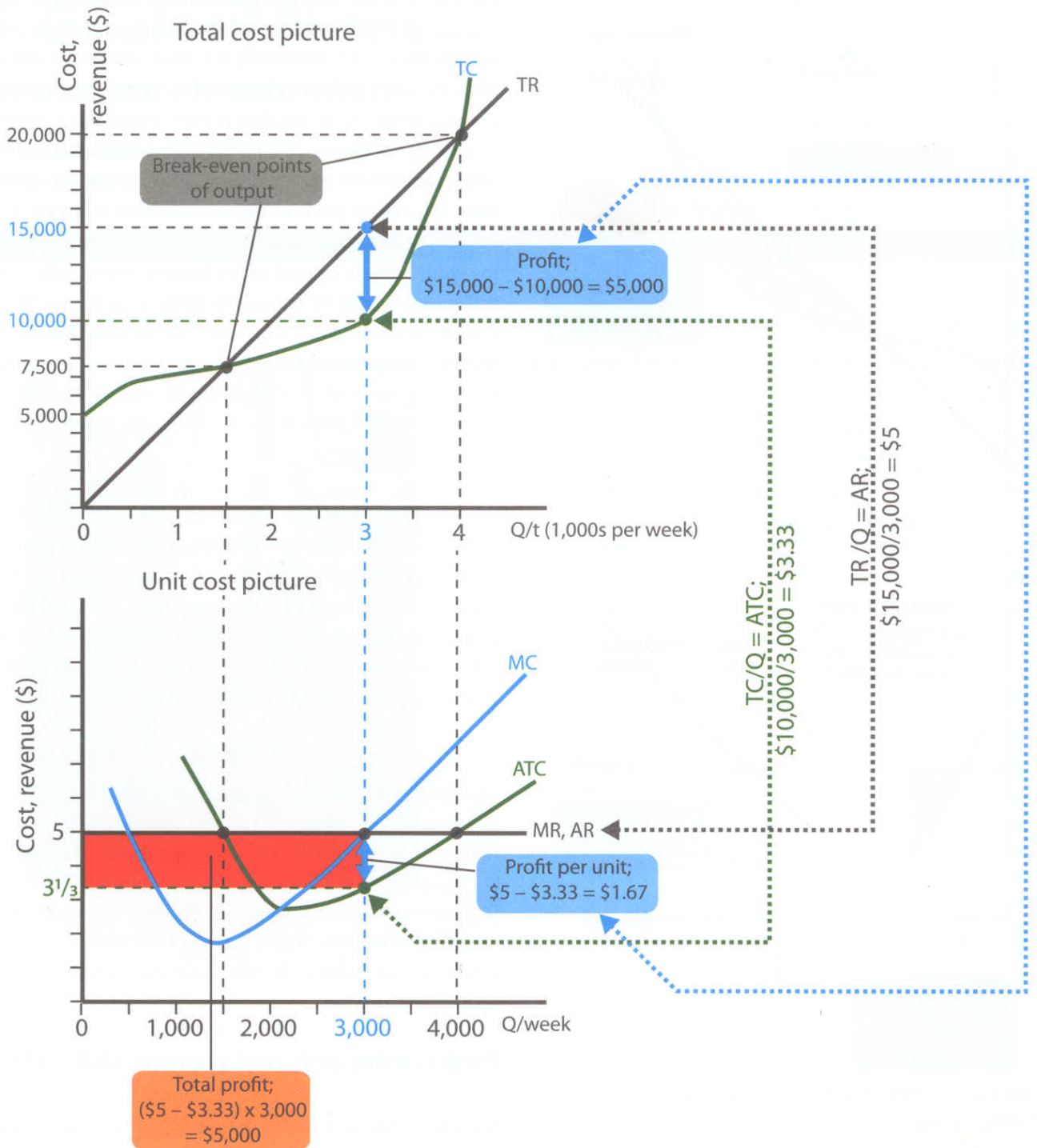
Figure 26.2 Profit maximisation using TR/TC and MR/MC curves

Figure 26.2 illustrates what happens when we add in an average total cost curve:

- The ATC curve will intercept the AR, MR curve at the same level of output that the TC curve intercepts the TR curve – at the *breakeven points* 1,500 and 4,000 units

- At a marginal cost of $5 and marginal revenue of $5 (profitmax!) the AR is $5 and the ATC is $3.33. The

'distance' between AR and ATC is $1.67 – this is the **profit per unit**.

- Total profit is the profit per unit times units sold, e.g. $1.67 x 3,000 which is $5,000 – the same as in the total cost picture of course.

You will be dealing primarily with the unit cost picture in the various market types to be covered in coming chapters. For the time being, just keep in mind the following two points:

1. **Finding profitmax level of output** is a matter of identifying the point where MC = MR.

2. **Finding profit** is a matter of deducting average total costs from average revenue and multiplying by total output. This is (AR − ATC) × Q.

Criticism and defence of the profitmax condition

Note that the model of marginal cost pricing is a purely theoretical rendition of how to optimise profits. Firms will in fact use other methods to set price, for example a cost-plus price strategy, where price is set at estimated average cost per unit whereupon a profit element is added on. Another method – increasingly common in, for example, the car industry – is to set a maximum price at the initial design stage of the product and then **build the product to conform to the costs**. Neither method corresponds to the MC = MR criteria and methodology outlined in this chapter, which makes it easy to criticise both the premises and predictive quality of the model as being unrealistic and abstract.

However, the model is not a 'blueprint' or 'roadmap' for firms to arrive at the best possible output level, but a scientific method of clarifying the concept of maximum possible profit. It is in fact more the reverse that must be true; if a firm is maximising profits, then it must be pricing at MC = MR! Our models in the theory of the firm are attempts to put into theoretical context what firms in all likelihood are doing. We are in effect trying to explain rather than prescribe. My friend Mats, a quite amazing biologist, once tried to explain to me how incredibly efficient ants are in their use of energy. He used any number of weird terms and models to prove his point and I was finally so confused that I tried to joke it off by asking, 'Yes, but do the ants know this?!'

'Probably not. But I don't explain it to them – I explain it to the students!'

Alternative goals of firms

We will often assume the firm to be a SR profit maximiser in our models. However, firms often behave quite differently in reality and forego SR profit for several reasons:

- One reason is that the firm might want to capture a larger share of the market and thus sets a relatively low price. Strategic **market entry pricing** is designed to maximise sales volume and attract and hopefully keep a loyal customer base over time, which could enable the firm to work up to economies of scale perhaps and thus ultimately make a profit. '**Predatory pricing**' is a type herein, where the (large!) firm tries to 'kill off' competition by setting a price that most competitors cannot match – the large firm can bear losses for a much longer period of time than small(-er) firms. This has been quite common in the market for air travel, where existing airlines have basically set loss-making prices to evict rivals from lucrative routes.

- Managerial theory puts forward the possibility that managers fall into a 'comfort zone' of profit and forego maximising profit in order to reach a reasonable target level of profit. This is known as **profit satisficing**[2] and implies that once a certain level of profit has been attained, entrepreneurs might look to attain other goals. A large multinational with millions of shareholders is not run by the owners but by professional managers – who might well act simply to keep shareholders happy by attaining targets and little more. I have often thought how this in fact applies to IB teachers in private schools; teachers' jobs are often dependent on achieving certain grade averages – if mock exams are finished and it is evident that the class average in economics will be well above the school's set standard, is there really any incentive for the economics teacher to put even greater effort into achieving an average of 7.0?! No, not really. And this line of reasoning has been shown to exist in several studies of both small and large enterprises. It should be noted with great care that the 'satisfactory' level of profit is going to vary greatly from firm to firm.

- Many times firms will act as '**good corporate citizens**' in price-setting. Pharmaceutical companies have been forced by some very strong public opinion to forgo excessive profits in order to fend off possible negative advertising effects and even boycotts. Firms must also increasingly act in accordance with *environmental concerns* and legislation; higher demands on firms via both consumers and interest groups can for example force firms to lower output and/or increase marginal costs to decrease pollutants.

---

2 No, I didn't misspell this. It's one of those terms which is a combination of two words, like stagflation (stagnation + inflation). Here the term is a combination of 'satisfy' and 'suffice' (to be sufficient or enough).

- Firms will often try to **keep prices stable** over time. Consumers do not like the insecurity and lack of predictability in not being able to plan future expenditure. Thus, increasing costs, changing exchange rates, or tax levies – to name but a few – would in many cases not affect a firms output level, in spite of a change in the MC = MR point.

- Selling at break-even price or even at a loss and still making an overall profit is quite possible when the use/application/upkeep of the good has a number of pricey complement goods in the future which are the sole proprietorship of the firm. Such **strategic pricing** is common for such goods as computer hardware which would need accompanying software or cars and service or spare parts. One of the more noticeable consumer goods of late to use this method is printers and ink cartridges, where manufacturers often sell the printer at below cost but then recoup losses on the sales of cartridges – which are increasingly protected by copyright laws in order to prohibit less costly versions produced by other companies. This has gone so far that producers have built in 'smart-chips' in order to hinder consumers from simply buying their own ink and refilling the cartridges. This has resulted in a series of complaints by consumer groups to the European Union which is currently investigating the case as being potentially harmful to competition.

- Finally, as we will cover later, the firm might want to **maximise revenue** instead of profit – this will give a different price and/or output than that of profit maximising.

The cynic in me has to add that no matter what the pricing strategy allegedly is, firms exist to make a profit. End of story. All the various goals listed above are simply ways for a firm to increase future profits.

POP QUIZ

More Revenue

1. Assume a profit maximising firm operating on a competitive market and producing at MC = MR. How would the firm react to an increase in MC? How would it react to an increase in MR?

2. Two competing soft-drink firms, Cepsi and Poke, are entering a new market where neither has been sold before. What type of pricing strategy would these firms consider?

Summary & revision

Profit maximisation is achieved by the firm when output is set where marginal costs equal marginal revenue. **Profitmax** is when MC = MR.

1. In the unit cost picture, profit is average revenue minus average total costs times quantity. Thus, **profit in the unit cost picture** is $(AR - ATC) \times Q$.

2. There are alternative goals for firms:

   a. **Market entry pricing:** setting the price at an attractive level in order to enter a market

   b. **Predatory pricing:** price is set at a level which competitors might not be able to match

   c. **Profit satisficing:** firms might look at other objectives once a minimum required level of profit is attained

   d. **Good corporate citizenry:** firms might forego maximum profit in order to be regarded as environmentally friendly or fair in terms of working conditions and pay package.

   e. **Price stability:** firms often forego profit maximum in order to keep prices stable over time

   f. **Strategic pricing:** setting lower prices on certain goods in order to gain higher profit margins on complementary goods

   g. **Revenue maximising:** firms can forego short run profits in order to instead maximise total revenue.

27. Perfect Competition

Key concepts:

- Assumptions of the perfectly competitive market model
- The firm as a price taker and short run profit maximiser
- The perfectly competitive market firm in the long run
- Shut down and break-even price
- Efficiency in the perfectly competitive market firm

'If a man writes a better book, preach a better sermon, or make a better mousetrap than his neighbour, tho' he build his house in the woods, the world will make a beaten path to his door.'[1]

Assumptions of the perfectly competitive market model

Recall that a poor model is one which does not do a good job of explaining or predicting reality. I would add that a good model is also one which is based upon realistic assumptions, meaning that the model is not too removed from reality by the time all the model's assumptions are in place. As we shall see, many of our models within this section turn out to be fraught with assumptions and stipulations which weaken their explanatory and predictive links to reality. Yet one should be careful in being distracted from the value of micro economic models that are abstractions of reality – they do indeed offer insight into most of the essential real-world features of economic activity in firms. The model of perfect competition is at one end of the extreme while monopoly is to be found at the other. We will go through them both, black and white, in order to mix a potpourri of grey – where in fact 99% of reality is to be found.

The model of perfect competition operates within the confines of five basic assumptions:

1. There are a **large number of firms** competing in the market: A single firm is powerless to influence total market output and/or price. Therefore the firm is a price-taker.

2. The firms are producing **homogeneous** (= identical) goods: Firm A's Widgets will in no way be differentiated from those of Firm B. This results in price competition.

3. Each firm is a **short run profit maximiser**: Firms operate under the profitmax condition of setting output where MC = MR.

4. There is **perfect knowledge/information** amongst both firms and consumers: Firms will have total knowledge of any improvements in technology and manufacturing processes, while consumers will be fully aware of all firms' prices.

5. There are **no barriers to entry**: Nothing hinders firms from entering the market in order to compete with existing producers. Such barriers could be insurmountably high initial (start-up) costs, lack of access to key technology or raw materials, and legal barriers such as not having necessary patent rights.

Comments on the assumptions

I would be lax in not commenting on the implicit (= underlying) necessity of having a legal framework and the rule of law for firms to operate within. This is one of the major obstacles in many developing countries, as the right to private ownership (often called **property rights**) and protection of these rights under a (democratic?) rule of law is often lacking. It has been duly noted for several years that Russia is in more dire need of lawyers than economists.

I would also be remiss in not offering some basic criticism of the assumptions. As a whole, I cannot think of one single industry or market that comes even close to fulfilling all of them. Very

---

1 Quote credited to Ralph Waldo Emerson (1803 – 1882)

few goods are in fact homogeneous – even water is branded and sold at ridiculous prices[2].

The closest one comes to homogeneous goods are probably to be found in basic commodities such as copper and potatoes. And looking at perfect knowledge, well, how many times have you bought something in a shop and then seen the same good in another at a lower price[3]?! All the assumptions above remove us from real life, yes, but again, our model serves us well as an approximation of real life. 'Perfect' does not mean 'best' or 'optimal' in any normative sense of the word but rather a situation where all factors of production can freely flow to the best possible use.

The firm as a price taker and short run profit maximiser

I often fall for the temptation to have a little fun in class here. I start out by drawing two identical supply and demand diagrams on the board and then I ask if they can spot the difference. Unfortunately they know me too well by then and I get a lot of smart-ass answers. Then I say that the one on the right shows the increase in market supply when one of the firms on the market doubles its output. 'Ahhhhh!', goes the class, as assumption number one (firm is a price-taker) occurs to them. So, if the firm cannot influence supply, yet is able to sell all its output at the given market price, what would market demand look like from the firm's point of view? You guessed it; demand facing the firm would be infinite.

The market for potatoes (Figure 27.1) shows that market supply and demand forces create an equilibrium price of 10 Danish Crowns (DKK). The firm cannot influence this on its own, and thus simply accepts that it is facing a fixed price. The demand curve for the firm is infinitely elastic, since the market can soak up any amount produced by any one supplier. This is the perfectly competitive market firm's demand curve, which looks confusingly like the average revenue (AR) and marginal revenue (MR) curve – which is because it is one and the same.
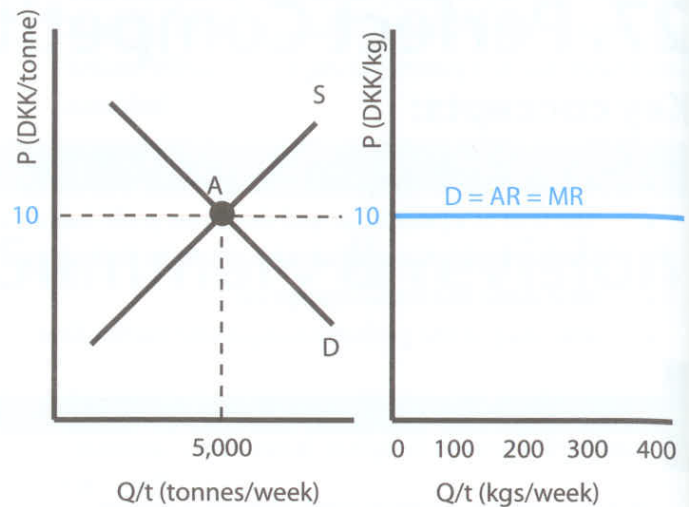

Figure 27.1   Demand and revenue for a perfectly competitive market firm

In summing up so far, we arrive at the following identity for the perfectly competitive market firm; the price will be the same as a horizontal demand curve, which is also the average and marginal revenue. Here's another credo for your presentation T-shirt to your economics teacher: 'Price equals murder!' Spelt '**P = MR, D, AR**'. Get it?

What about the supply curve for the firm? Well, consider that the firm will always set output where marginal costs equal marginal revenue and that the demand curve is given by market supply and demand. A change in, say, demand, would shift the demand curve for the firm also – and the MC = MR point would change. Figure 27.2 shows that when market demand increases from D_0 to D_1 and decreases to D_2, the demand curve (which is also the MR and AR curve) for the firm shifts upwards, *along* the upward-sloping MC curve! Any change in MR will change the profitmax intersection of MC = MR. The MC curve is thus the firm's supply curve and is the ultimate reason why market supply curves are upward sloping.

---

2   I just saw a report on TV from a wealthy area in the US where the price of bottled water at a rather upmarket store was far higher than the price of petrol.

3   It's embarrassing, but I must tell this story. For five years I had been going into the school's student café for my daily dose of chocolate. One day a student of mine asked me why I didn't buy the chocolate from the vending machine which for five years had been parked right outside the entrance to the café and had the same chocolate bars at a 20% lower price. I think this was the most embarrassing event in my life which didn't involve a woman or a head waiter.
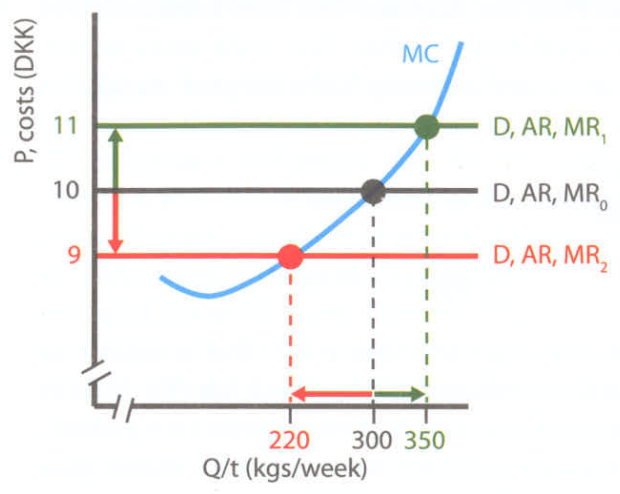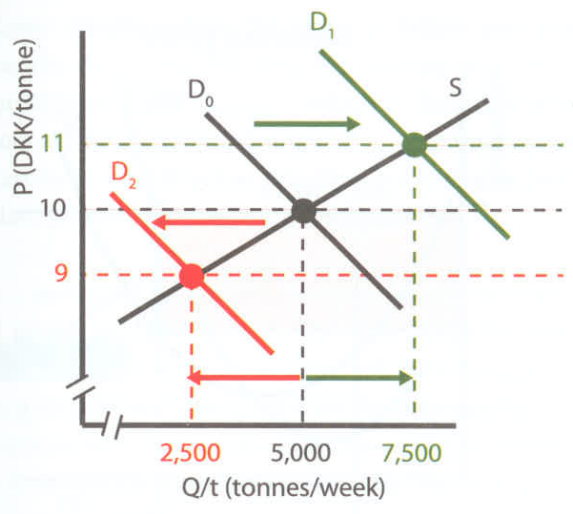
Figure 27.2 The PCM firm's supply curve

I will elaborate on the short and long run supply curve for the firm in the chapter on break-even and shut-down points further on. I introduce the concept at this stage in order for you to be able to adequately follow the examples on increasing and decreasing output levels in a firm.

A Little Depth

PED seen from a firm

Perfectly competitive firm as price-taker

Consider the following: Perfo-Firm is one out of 4,000 competing firms in a market for Widgets; PED is 1; total output is 60 million units; the market price of a Widget is 10 Swedish Crowns (SEK); and each of the 4,000 firms has an equal share of the market.

How would the market be affected by a 100% increase in output by Perfo-Firm?! Lets look at the market change first and then try to show why Perfo-Firm is a price taker.

Here are the main figures:

Total market output = 60 million. Average output of a single firm, e.g. Perfo-Firm:

$$=> \frac{60\ 000\ 000}{4\ 000}$$

$$= 15,000 \text{ units.}$$

Market price = 10 SEK. PED = –1. Percentage increase in total market output due to 100% increase by Perfo-Firm:

$$=> \frac{15\ 000}{60\ 000\ 000}$$

$$= 0{,}00025\%.$$

Total impact on market: remember, we know the PED of Widgets, and that:

$$PED = \frac{\%\Delta Q_d}{\%\Delta P}$$

How would the market price be affected by Perfo-Firm's doubling of output? As we know the PED and increase in quantity of the Widgets on the market, it's a matter of working out the change in price:

$$-1 = \frac{+0.00025}{P}.$$

We would see the market price fall by 0,00025%, or 1 SEK × (1 – 0.0000025), which gives a new market price of 9.999975.

Conclusion: When a single firm doubles its output, the market price falls by 0.000025 SEK! (About 0.003125 US cents at the current exchange rate.)

HIGER LEVEL

MICROECONOMICS

Perfo-Firm's PED; this individual firm faces a demand curve which gives a different PED than that of the entire market. When the firm raised output by 100%, the price changed by -0.00025%.

$$PED = \frac{\%\Delta Q_d}{\%\Delta P} \Rightarrow \text{PED for Perfo-Firm is:}$$

$$\frac{+100}{-0.00025} = -400,000!$$

Conclusion: The single firm faces a PED that is 400,000, or the functional equivalent of infinity. That is why the D-curve for the firm is horizontal for the firm operation on a perfectly competitive market – the firm cannot affect the market price and is thus a price taker.

The perfectly competitive market firm in the short run and long run

Short run price and output

The perfectly competitive market firm ('PCM firm' henceforth) is basically left with two decisions in the short run; whether to produce and how much to produce. We will start by assuming that the firm has answered 'Yes' to the former and is now regarding the latter. In line with our assumption of profitmax, the PCM firm will set output at the point where MC equals MR. Being a price-taker, the price is set by market forces (supply and demand) and the firm will have three possible outcomes in the short run, as shown in Figure 27.3 below. (Note: AC is henceforth used for average total costs unless otherwise stated.)

Normal profit

Abnormal profit

Loss
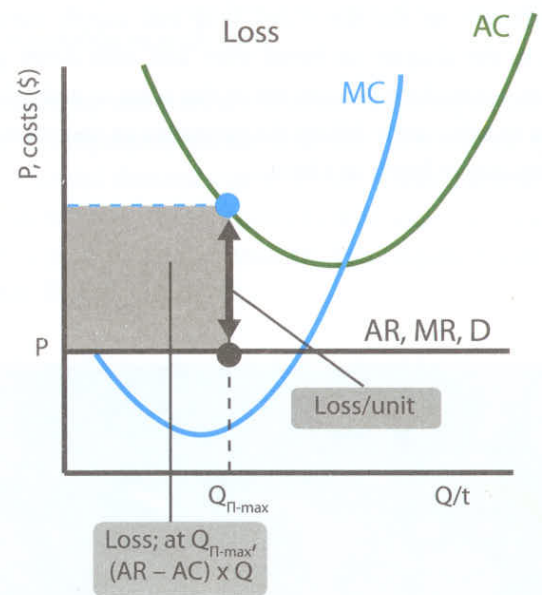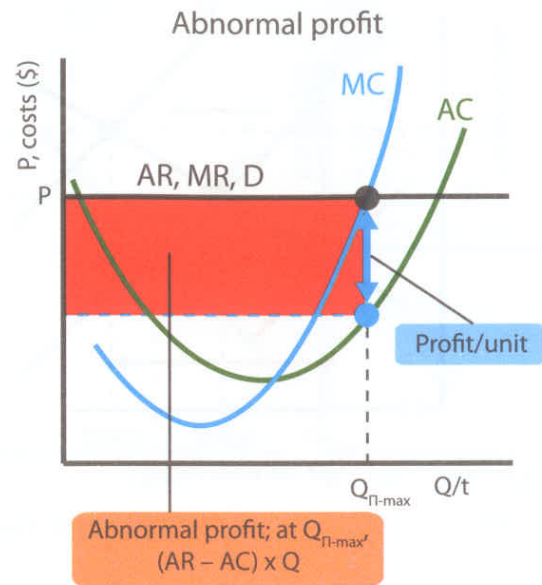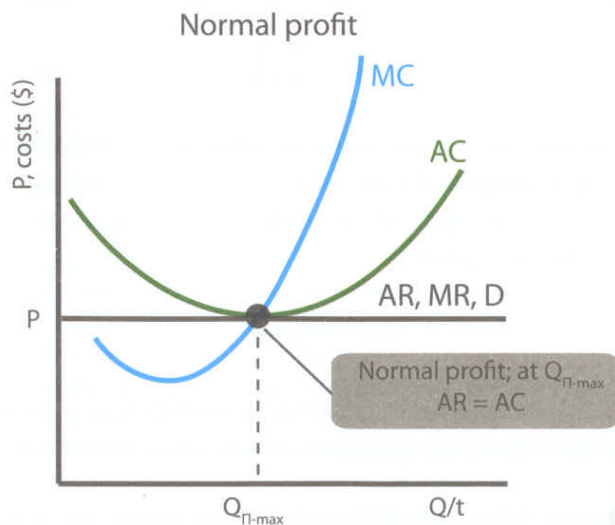
Figure 27.3 Short run possibilities in a PCM firm

Recall that profit is the result of subtracting total costs from total revenue, which in the unit-cost picture corresponds to subtracting average costs from average revenue. Figure 27.3 shows three possibilities:

- **Normal profit:** When the market price equals the AC of the PCM firm, the firm will *break-even*, i.e. it will enjoy normal profits. This is shown in the diagram to the far left above, as the MR = MC point coincides with average costs. As AR = AC there is a normal profit.

- **Abnormal/supernormal profit:** The middle diagram illustrates a situation where the market price (and thus the MR, AR curve) is above the average cost. The firm sets output at $Q_{\Pi\text{-max}}$ and earns an *abnormal profit*, shown by the red area.

- **Loss:** Finally, when the price is below any point on the AC-curve, the firm will operate at a loss, as profitmax output ($Q_{\Pi\text{-max}}$ – which is the same as the loss minimising level of output; $Q_{\text{loss-min}}$ in the diagram) results in an AR below AC. The loss is shown by the dark grey area in the far right diagram[4].

A firm in a perfectly competitive environment can only enjoy abnormal profits in the short run. The same holds for losses, which makes intuitive sense as no firm will be willing/able to uphold long term losses.[5] The market mechanism together with the assumptions will act to create **long run equilibrium where the PCM firm will earn normal profits only**. We look at both scenarios in turn.

---

4   Students often ask why the firm continues to set output at MC = MR even when average costs are higher than the price. The answer is that this point is also the *loss minimising* point of output, i.e. the firm would make an even greater loss by setting output at any other level than the MC = MR point.

5   Possible exceptions; US car manufacturers. Look up the USD80 billion dollar US government bailout of GM and Chrysler during 2008.

Short Run abnormal profit – market entry

I finally got full colour to play around with for this edition! I want you to think of the red area of abnormal profit in Figure 27.4 below as a bullfighter's cape. You see, a firm which is making a profit above and beyond possible alternatives – an abnormal profit – would be to other firms what the cape is to the bull; a signal to charge in. The firm depicted in the diagram on the left has an AC curve where the AC_{min} point is below the market price, P_0. The firm sets output at the profitmax point of MC = MR and thus has an average revenue which is above average cost. This is the abnormal profit per unit, shown by the double-edged arrow, and this times the quantity shows the total abnormal profit for the PCM firm – the red rectangle.

What then happens, keeping in mind the assumptions of 1) *free market entry* and 2) *perfect knowledge/information*, is that new firms will be attracted and of course enter the market. This increases supply from S_0 to S_1 causing the price to fall from P_0 to P_1. Falling market price will lower AR for the single PCM firm, creating a long run equilibrium where once again AR = AC. The firm's short run profit is thus eroded in the long run by market entry.

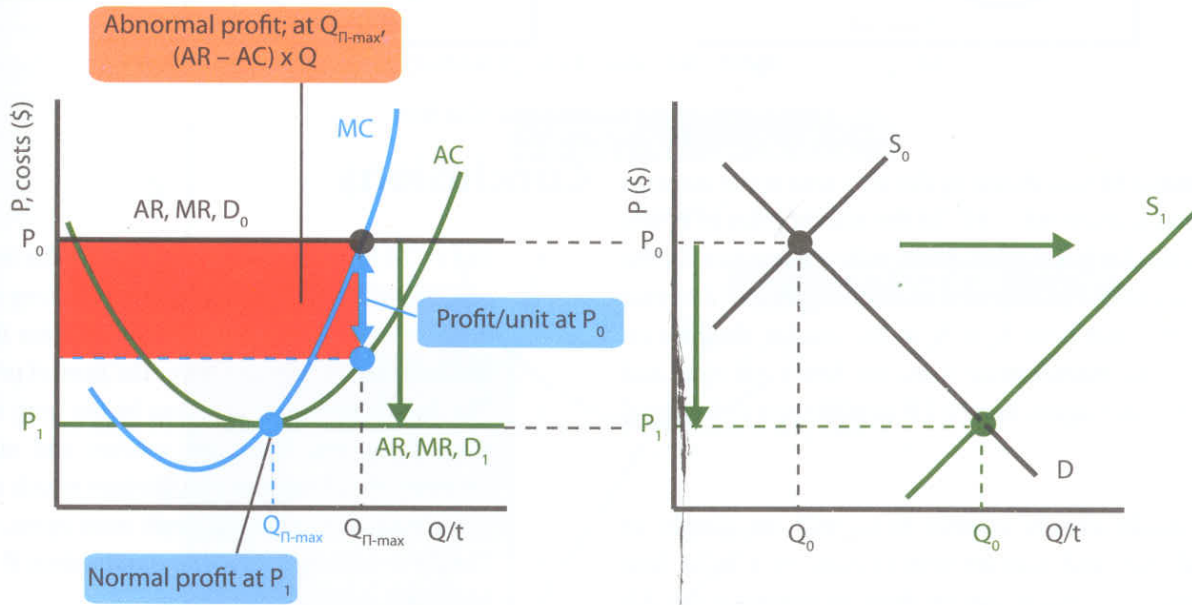
Figure 27.4   Short run **profit** in a perfectly competitive market firm

QUESTION:

How is it possible that the individual firm in Figure 27.4 above is producing less, while market output has evidently increased?!

Short Run Loss

Assume that firms which have been attracted to the market (as outlined in the example above) increased market output to the extent where the increase in supply lowered the market price to a level where individual firms made losses. This is the situation shown in Figure 27.5 for a loss-making firm. At a market price of P0, the firm's AR is below AC. The firm will still produce at MC = MR ('loss-minimising' point in this case) and will run at a loss, shown by the double-sided arrow. Total loss is the grey rectangle. As firms begin to exit the market over time – switching to more attractive producer substitutes – the market supply curve will shift to the left. Firms, such as the one in our example, which have managed to 'ride out the storm', will see how the price once again rises to a level where a normal profit can be made; where AR = AC.
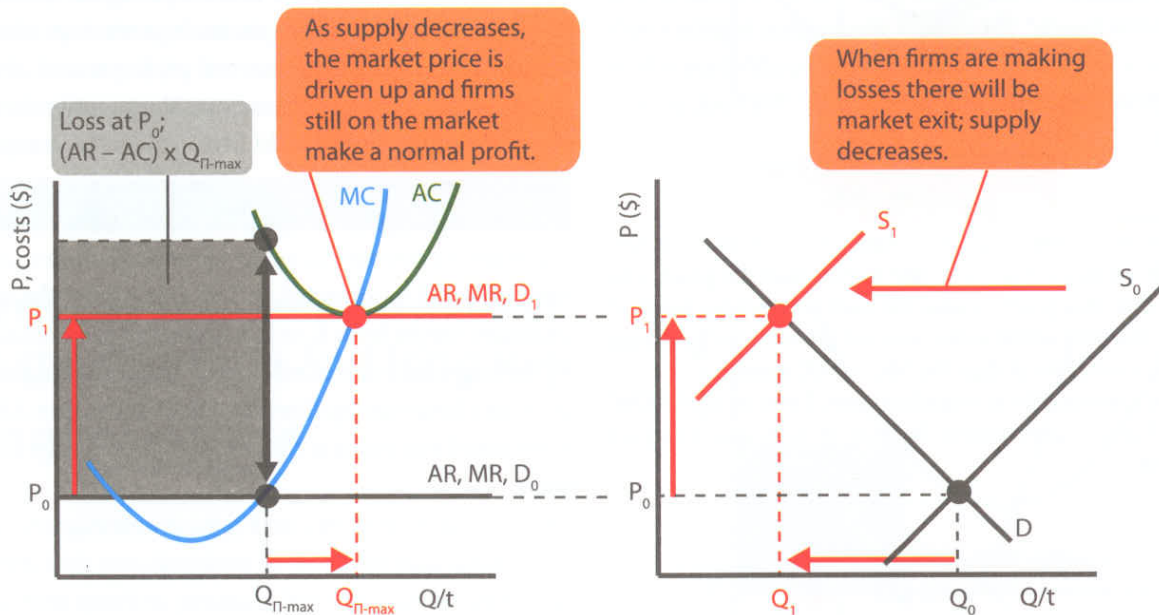
Figure 27.5   Short run **loss** in a PCM firm

Conclusions

The firm makes a SR loss shown by the grey area in the diagram on the left; loss is when AR < AC. At the market price of **P₀** this firm, and in all likelihood other firms, will run at a loss. In the LR, firms will exit the market and market supply will decrease – shown by the shift from **S₀** to **S₁** in the market diagram on the right. As the market price rises, the firm's AR rises and when AR = AC once again, there is **LR equilibrium** and normal profits.

As you no doubt already suspect, firms will not simply sit with crossed arms and wait for better times. As a PCM firm is a price taker, not much can be done to influence the AR side of the coin, so firms are focused on lowering costs. A firm running at a loss will have to find ways to become more efficient (i.e. lower MC) and/or decrease costs in general. One of the most common methods used to decrease costs is to decrease the amount of labour used in production and to try to use remaining labour more efficiently. This is a major part of the issue of unemployment and the business cycle dealt with in Section 3.5.

- The PCM firm can make abnormal profits in the short run, but will make a *normal profit in the long run* as lack of entry barriers allows new firms to enter the market and increase supply and lower the market price.

- The firm cannot run at a loss in the long run either since firms will leave the market and supply will convene on a long run equilibrium which allows the (surviving) firm a normal profit once again.

- The *LR equilibrium level of output* is thus: $P = AC_{min} = MC = AR = MR$.

Shut down and break-even price

"I'll be here 'til the end of time
So you gotta let me know
Should I stay or should I go?" (Awesome song by The Clash, from 'Combat Rock', 1981)

When a firm is earning normal profit output is at the point where AR = AC, this is the **break-even point** of output. What then, is the point where a firm will leave the market? Simple,

the shut-down point is the output level where it is equally costly for the firm to continue producing as it is for the firm to leave the market. If the firm can cover all of the variable costs and at least some of the fixed costs then it has an incentive to remain on the market. If the price falls below the AVC then the firm will not even cover the variable costs – and leave the market. Hence, the point where the firm must decide whether to remain in the market or leave is when AR = AVC. This is the **shut-down point.**

I will use figures to show this, as my experience is that it is easier to comprehend actual numbers rather than using 'points A, B and C'. The firm depicted in Figure 27.6 is subjected to ever-lower market price. Assume that the original demand on the market gives a market price of €10, which is the firm's MR and AR. This is the long run equilibrium and also the break-even point, as the firm covers all its costs – even opportunity costs – earning it a normal profit.

Say that for some reason (either increasing supply or decreasing demand) the market price starts to fall and subsequently the firm's AR, MR, D-curve falls to a price level of €6. Being a profit maximiser, the firm sets output where MC = MR, which is now at 80 units rather that 100. At this output level the firm cannot cover all its costs; ATC at an output of 80 is €11. The firm loses €5 on each unit produced, giving an overall loss of €400 (€5 × 80 units).

> **Definition: 'Break-even price and shut-down price'**
>
> When a firm is selling at a price where total revenue is equal to total costs one speaks of the break-even price. In the unit cost picture this is when P (AR) = AC.
>
> When a firm is operating at a price level where the price only covers average variable costs and none of the average fixed costs, then the firm must make a decision whether to remain on the market or exit. This **shut-down price** is defined as P = AVC.

Note: TFC = (ATC − AVC) x Q. At an output of 70, we get (€12 − €4) x 70 = €560.

Figure 27.6 Short run loss in a perfectly competitive market firm

Why doesn't the firm leave the market at a price level of €6? Consider the choices facing the firm:

I.   Stay in the business and make a loss of €400

II.   Leave the business and make a loss of €560, which is the total fixed cost (see TFC calculation in Figure 27.6 above)

This is not much of a choice, rather a lack of options. The firm will have a strong incentive to stay on the market during the short run, hoping perhaps that either market price will increase or that increased efficiency and/or cost-cutting can lower MC and AC to a normal profit level again.

If, however, the market price falls even further, to €4, then the options become:

I.   Stay in the business and make a loss of €560

II.   Leave the business and make a loss of €560

This is what my grandmother called 'Choosing between sleeping on a rock or a hard place'; i.e. not much of a choice. The firm's TR (€4 × 70 = €280) will be identical to the TVC – which means that there is no contribution towards covering the fixed costs. The firm is making a loss of €560 by staying in the business and would make the same loss by leaving it. The point where P (AR) = AVC is therefore the shut-down point for the firm. The firm will not produce at a lower price level – just consider the options at a price of €2. The firm's TR would be €120 (€2 × 60) and TC would be €720 (€12 × 60) leading to a loss of €600. The reason is that at a price (e.g. AR) of €2 the firm would not even be covering all its variable costs so total costs would be greater than total fixed costs alone. At any price below AVC the firm will leave the market.

Conclusion: As long as the firm has an AR above AVC, the firm covers variable costs and at least some of the fixed costs – therefore there is an incentive to stay in the business in the short run. **The shut-down point is when P (AR) = AVC.**[6]

SR and LR supply curves for the firm

At any price (AR) above €4, the AR received will help to cover at least some of the fixed costs. This contribution to covering total costs would enable the firm to stay in the business during the short run, as any price which is above AVC will mean less loss to the firm than shutting down and still having to pay the total fixed costs.

- The PCM firm's supply curve in the *short run* is thus the portion of the MC curve which is *above the AVC curve.*
- The *long run* supply would be the portion of the MC curve *above ATC* as no firm could withstand indefinite losses and would ultimately have to leave the market if it did not cover all costs.

Back to basics; explaining the supply curve

I finish this chapter by 'circling around' and connecting the firm's supply curve to the market supply curve. Building on the proposition that firms will supply at any possible level above the average variable cost curve in the short run, the market supply can be derived from this. Figure 27.6 assumes a market of three

6   The father of one of my students was the President of the Mexican branch of one of the world's largest steel producers. He kindly invited me to his rather select club one Saturday morning for breakfast. During our talks, I probed him on the issue of 'shut-down' at his operations during the on-going financial crisis (December 2008). Steel is very sensitive to cyclical changes in demand and the impending recession was of course causing steel prices to plummet world wide. His answer was very clear: "As soon as I cannot get a price that justifies production – in other words, does not cover production costs – I close the plant down!" So much for a 'short run supply-curve'.
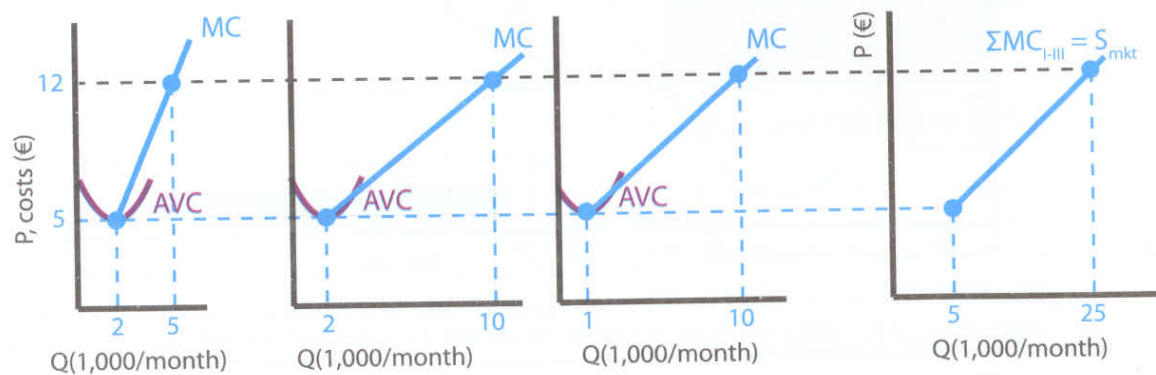

Figure 27.7   Foundation of market supply; Σ MCPCM firms = market supply

firms (it wouldn't matter if I used 35,000 firms, but my editor has told me to save space) which all have different marginal and average costs. Only the portion of the MC-curve above AVC is of interest in this case, and summing the individual firms' output at various prices yields the short run industry/market supply curve; 5,000 units per month at a price of €5 up to 25,000 units per month at a price of €12.

This is when most of my students have something of an 'AH-HA!' experience. The concept of increasing marginal costs coupled to the profitmax condition of MC = MR renders each individual firm's supply curve, and the aggregate thereof is market supply. It helps in understanding that when firms' marginal costs are affected by technology, production improvements, lower costs of labour and raw material etc, the total market supply will change as a result.

Efficiency in the perfectly competitive market firm

I had this brilliant idea once for 'Floor-laying Instructions For Complete Morons'. I suggested to Glenn, my friend in the house-renovation business, that we could market easy-do-it-yourself kits for people redoing their bathrooms. On the bottom of each floor tile we would print 'If you can read this, you are doing it wrong!' There are two versions of efficiency for economists; **productive efficiency** and **allocative efficiency**. Productive efficiency means 'doing things right', while allocative efficiency means 'doing the right things'. Putting the tiles in with the least amount of resource use, i.e. doing it correctly, is productively efficient. Putting in the kind of tiles that your better half (= wife or husband) likes the most is allocatively efficient. Hmmm, I suppose I should be a bit more technical…

Productive efficiency – doing things right

A firm is *productively* efficient when total use of resources (factor inputs) results in the lowest possible cost per unit of output. This would be the point where average total cost is minimised. Any other level of average costs would be sub-optimal. A perfectly competitive industry is characterised by forever attempting to implement production methods and technology that increases efficiency – see the LRAC curve.

Allocative (economic) efficiency – doing the right things

If the market is producing 1,000 red pens at the lowest possible average cost, but market demand enables it to sell only 800 then the firm is utilising resources wrongly. There is waste since the

resources used in the production of the excess 200 units could be better allocated. Firms could use the resources to produce goods which are in higher demand, say green pens. Thus, on a market level, optimal *allocative* efficiency (also known as economic efficiency) occurs when supply equals demand on the market.

In regards to individual firms, the definition of allocative efficiency is that the individual firm is producing the correct quantity of the right goods – 'doing the right things'. In stating that the correct quantity is produced we are in fact saying that since the last unit produced costs exactly what the consumer is willing to pay, resources have been optimally allocated. As consumers buy 'on the margin' (see Section 2.2) then the price paid reflects the value placed on the good. Resources have been optimally allocated when there is no waste, i.e. when the price equals marginal cost of the last unit produced. This occurs when the output level is where P (AR) = MC for the firm. When all firms fulfil this criterion then supply equals demand on the market.

> **Definition: 'Productive efficiency'**
>
> **Productive efficiency** occurs when goods are produced at the lowest possible cost per unit, taking into account all costs arising. For a firm, this is when output level is at AC_{min}.
>
> **Definition: 'Allocative (economic) efficiency'**
>
> When there is no waste in resources in the production of goods, allocative efficiency is optimised. This point is reached when there is zero excess supply or demand on the market and the marginal revenue is equal to the firm's marginal cost. S = D…and… P (AR) = MC.

The firm operating within a perfectly competitive market will be both productively and allocatively efficient in the long run. It has been shown that the firm cannot have an abnormal profit in the long run due to the entry of new firms whereby the subsequent increase in supply and lower market price will dissolve any such profits. Nor can the firm survive endless losses. Figure 27.7 shows the LR equilibrium for a PCM firm; output is at $P = AC_{min} = MC = AR = MR$.

- **Productive efficiency:** The LR equilibrium for the perfectly competitive market shows that AR = ACmin. The firm is productively efficient.

- **Allocative efficiency:** The horizontal demand curve will set output along the upward sloping MC curve, inevitably forcing the firm to produce where the marginal revenue equals the marginal cost. In LR equilibrium, P (AR) = MC. The firm is allocatively efficienct.

PCM firm

> **Productive efficiency:** PCM firm produces where AC is minimised.
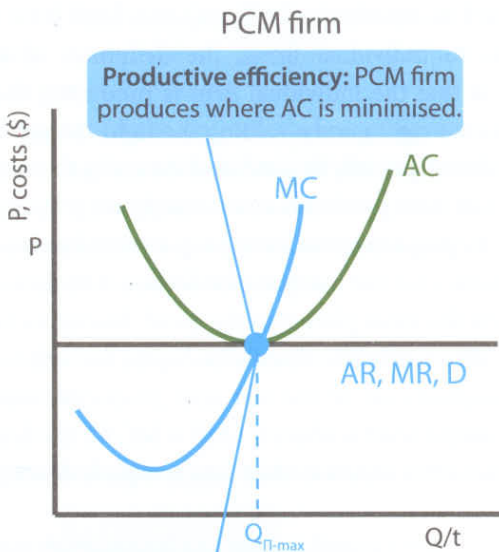
> **Allocative efficiency:** The PCM firm produces where P (AR) = MC. Market equilibrium is attained as no excess supply or demand exists.
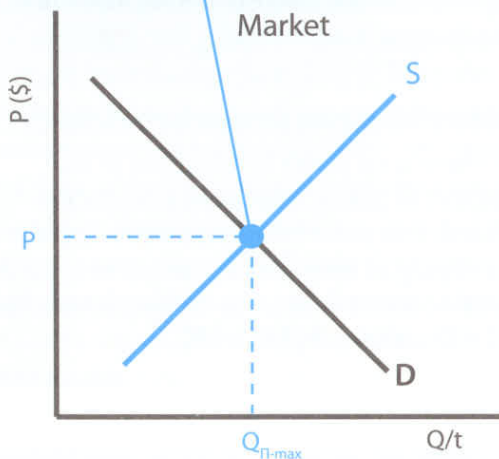
Market

Figure 27.8 Short run loss in a perfectly competitive market firm

In summarising the issues in perfect competition and efficiency, we must conclude that the PCM firm in the long run produces at an output level where **P = AC$_{min}$ = MC = AR = MR**. This identity fulfils the criteria for both productive and allocative efficiency, i.e. the model shows that the perfectly competitive firm (and market) is both productively and allocatively efficient in the long run. Yet, as seen in Chapters 16 – 21 on market failure, there are a good many 'howevers' involved in the real world.

Summary of how to 'interpret' cost curves

The curves used in *theory of the firm* tell us different things about the state of the firm's environment and possible choices. The main ones are:

Profitmax output; Firm's output is where MC = MR

Profit; TR - TC and AR - AC (note that opportunity costs are included)

Productive efficiency; Firm's output is at AC$_{min}$ (AC = MC)

Allocative efficiency; Firm's output is where P (AR) = MC (also when S = D on the competitive market)

A Little Depth

New technology and the perfectly competitive market

An estimate shows that global GDP per capita (in constant purchasing power dollars, 1990s) was around $400 for most of the last 2,000 years. In 1800 this jumped to around $700 and by the year 2000, global GDP per capita was over $6,000. During the same period, England's GDP/capita increased 20-fold and the USA's 36 fold. It doesn't take an intellectual quantum-leap to understand why this enormous increase in productivity occurred during the end of the 1800s; the industrial/technological revolution.

A central point in any competitive environment is the ability to make better use of resources. A firm which comes up with a better method of production or implements new technology at an early stage will be able to increase output per unit of factor use. This is equivalent to saying that the cost per Widget falls. The first diagram in Figure 27.9 portrays a firm which has implemented new technology, for example a computer-assisted manufacturing system. This lowers the marginal cost (from MC$_0$ to MC$_1$) and also average costs (AC$_0$ to AC$_1$) – which is in line with previous diagrams of increased capital use lowering AC in the long run. The firm will increase output to Q$_1$ in line with the MC = MR condition of profitmax. (Note that the firm is producing beyond AC$_{min}$.) The firm's increase in output does not influence market output or price at P$_0$ and Q'$_1$. The firm is now able to enjoy an abnormal profit (the red area) as AR is higher than AC.

PCM firm

Abnormal profit MC_0

The increase in market supply decreases AR, MR, D

AC_0

MC_1

AR, MR, D_0

AC_1

AR, MR, D_1

Normal profit at P_1

Q_0 Q_2 Q_1 Q/t

Market

Market entry plus adopting new technology causes an increase in supply

S_0

S_1

Q'_0 Q'_1 Q/t

Figure 27.9 New technology and SR abnormal profit

As the firm is operating in a competitive market, two key assumptions of the perfectly competitive market come in to play and cause market forces to react and increase market output;

1) Perfect knowledge/information – other firms on the market will know about the new technology and start to adopt it;

2) No barriers to entry – there will be an incentive for other firms to enter the market as abnormal profits are being made.

Taken together, market output will increase, shown in the lower diagram by the shift of S_0 to S_1. As more firms adopt/enter causing an increase in supply, output increases from Q'_0 to Q'_1 and the market price falls from P_0 to P_1.

The lower market price accordingly shifts the AR, MR, D curve downwards to AR, MR, D_1, in the lower diagram. This will lower the firm's output as MR moves along the MC curve to where long run equilibrium is attained at Q_2. Once again, at P_1 and Q_2, the PCM firm is making a normal profit, as abnormal profits have been eaten away by existing firms' technology adoption and additional firms entering the market. Conclusion: The 'path' of output increase in the long run here is in total accordance with the LRAC curve done in Chapter 24.

*Delong, J. Bradford 2000a Estimating world GDP, one million B.C. – present. www://econ161.berkely.edu/tceh/2000/world_gdp/estimatina_world_gdp.html

A little rumination… 'perfect'…

A perfectly competitive market could easily be considered a dream world stretch of the imagination - 'Mermaidomics' to coin a phrase. There *are* barriers to entry such as high initial costs and legislative barriers; there is *no such thing* as perfect knowledge/information since no one person could possibly amass all prices for example[7]; many markets have only a *few* suppliers; and many firms *do* have the power to set price and/or quantity.

However, our model will be broadly in tune with reality as many firms will in fact be reaching for solutions that increase profit – they will act to lower costs, enhance efficiency and increase output. The competitive forces on a market do create a dynamic over time where firms tend towards efficient solutions in order to reap profits.

In the long run, it is empirically (= practically observable) apparent that many firms will be struggling to make the most of scarce factors; this is a competitive outcome leading to efficiency. The theory of the competitive market can help us to understand this.

1. Illustrate and explain the concept of allocative and productive efficiency in a perfectly competitive market.

---

7 Yes colleagues, this is debatable. Currency markets might be considered to be perfectly symmetric in terms of available information. Yet I *always* ask myself why there are willing sellers of, say, US dollars no matter what…
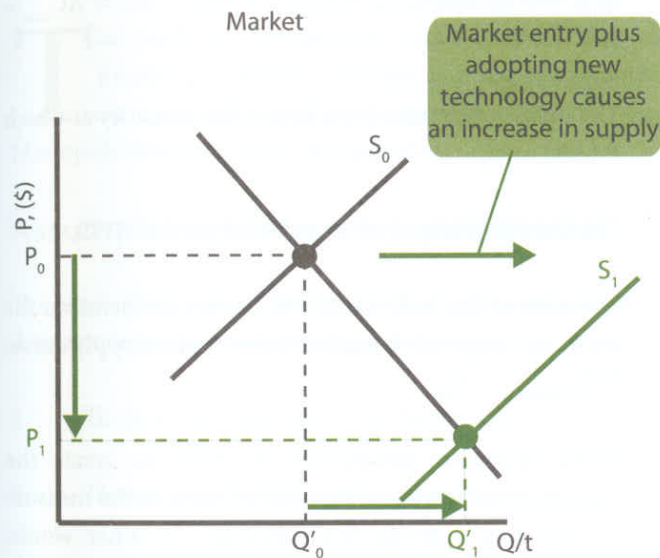
2. Why does our assumption of each individual firm being too small to influence the market have such an effect on the firm's demand curve?

3. Explain why the sum of individual firms' MC curves is the market supply curve.

4. How would increased labour costs affect the PCM firm's unit cost picture? Explain. Use appropriate assumptions!

5. Why can a PCM firm only make a normal profit in the LR according to our model?

6. A firm is operating at a loss. Explain why the firm might stay rather than exit the market.

7. Assume a total of three firms all operating in equilibrium on a perfectly competitive market. AR = €8. Firm A can produce a quantity of 6,000 at a MC of €8; Firm B can produce 4,000 at MC = €8; Firm C can produce 2,000 at AC_{min}. What is total market demand?

8. Using diagrams, explain the probable effects of a permanent increase in demand on a perfectly competitive market. Might the price actually decrease ultimately?

9. Tricky one (and outside the syllabus!): How would a lump-sum tax (which is a one-off or 'once only' tax – for example, if all firms had to pay £1,000 pounds as a 'one-only environmental fee') affect MC and AC?

Summary and revision

1. The perfectly competitive market model is based on **five main assumptions**; 1) a large number of firms; 2) the firms are producing homogenous goods; 3) the individual firm is a short run profit maximiser; 4) there is perfect knowledge and information; 5) there are no barriers to entry.

2. The assumptions result in each individual firm on the perfectly competitive market being a **price taker**. Thus; $P = AR = MR = D_{firm}$ ("Price is murder").

3. The **profit maximising level of output is where** MC = MR. The firm's supply curve is the marginal cost curve since any change in the market prices will move the profitmax point of output 'along' the upward sloping portion of the MC curve.

4. **Profit** is shown by $(AR - AC) \times Q$.

5. In the **short run**, the PCM firm can be making an abnormal profit (AR > AC); a normal profit (AR = AC); loss (AR < AC).

6. In the long run the PCM firm will make a **normal profit**. Abnormal profits will attract market entry and the price will fall…while losses will cause market exit and the market price will rise. Hence, in the long run, the PCM firm will be operating where $P = MC = AR = AC_{min} = AR$.

7. The **break-even** price for a firm is the quantity at which P (AR) = AC.

8. The **shut-down** level of output is where P = AVC.

9. The **sum of the individual MC curves** for firms on the perfectly competitive market is the market supply curve. $\Sigma MC_{individual} = S_{mkt}$

10. Using factors of production in order to create the highest possible output per unit of input is the measure of efficiency – **productive efficiency**. In other words, when goods are produced at the lowest possible cost per unit there is optimal allocative efficiency. This is where $Q = AC_{min}$. The issue of efficiency is basically 'Firms are doing things right!'

11. The question of whether consumers are then willing and able to buy the amount of goods produced is also a measure of efficiency – **allocative efficiency**. When firms are producing goods in the correct quantity – as defined by market demand – then there is no excess in supply or demand and firms are allocating factors of production allocatively efficiently. This is when output is set where P = MC. The issue of efficiency is basically 'Firms are doing the right things!'

28. Monopoly

Key concepts:

- Assumptions of the monopoly model
- Barriers to entry
- Revenue and costs in the unit cost picture
- Profit maximising price and output
- Revenue maximising price and output

I think it's wrong that only one company makes the game Monopoly. Steven Wright

The perfectly competitive market model is a rather theoretical extreme of firms as price takers operating on a market where no long run profits are possible. Monopoly firms are quite the opposite!

Assumptions of the monopoly model

The assumptions of the monopoly market model are markedly different from the competitive model:

1. There is *one firm* supplying the market.

2. There are very high *barriers to entry*.

One assumption, however, remains the same:

3. The firm is a short run *profit maximiser*, just like the perfectly competitive market firm.

In fulfilling the assumptions above, a monopoly firm will have a great deal of power on the market, often at the expense of the consumer. A single monopoly firm is able to choose to produce at any point along the market demand curve. Therefore the firm is a *price-setter*. This power is a consequence of the lack of available consumer substitutes. A monopoly must – per definition – be supplying a good which is quite unique in terms of perceived utility, and therefore cannot be replaced. Postal service has long been an example of monopoly power, but (as seems to be the case in the long run) this power has been eroded over time as private delivery firms and email increasingly encroach this market.

Barriers to entry

Firms attempting to enter a market could experience **technical barriers**, where the monopoly firm enjoys decreasing marginal costs over the span of market demand (see **natural monopolies**, below), for example special techniques and production methods involved in production, or ownership of unique factors of production (e.g. raw material). An example of the former would be Pilkington Glass of England which in the early 1950s invented a revolutionary process for producing sheet glass used in windows and then licensed the technology to firms around the world. As for **controlling a vital factor** input, ALCOA (Aluminum Company of America) of USA controlled most of the world's bauxite (from which aluminium is produced) until the 1940s.

There could also be **legal barriers to entry**, for example intellectual property rights such as patents and copyrights. The firm could also enjoy government-granted production rights in public utilities such as postal service and telecommunication. Microsoft Corporation has a monopoly on both DOS and Windows and many countries will have legislation in place governing which companies are allowed to supply electrical power and telecom cables to the market.[1]

It is also quite common – even likely, according to some studies – that firms which are able to will attempt to **create barriers** in the aims of constructing a monopoly-like market situation. The monopoly firm's activities on the market could dissuade other firms from entering the market. A monopoly might engage in **predatory pricing**, whereby the firm seeking to uphold monopoly power sets the price lower than smaller rivals and potential entrants – see Chapter 32. Additionally, a monopoly

---

1 Have you noticed that all Microsoft product names are accepted by the automatic spell-checker in Word?! Just try writing 'MSWord' and then 'MSNerd'!

firm could aim to control the supply on the market by buying up the output of other firms in order to control end-user supply. The diamond merchant de Beers previously controlled – via a **cartel** – some 70% of the market for diamonds (depending, once again, on how the market is defined) which it has consistently tried to uphold by controlling not only the mines but also the wholesaling chain via purchasing and stockpiling diamonds from mines not under company control.

A final key barrier to entry is the existence of economies of scale. An incumbent (= existing) firm enjoying large benefits of scale will dissuade potential entrants simply by having far lower average costs than any newcomer can attain. This often creates a so-called **natural monopoly** which will be looked at in Chapter 29.

Revenue and costs in the unit cost picture

Picture the market demand curve. Now cross out 'Market demand' and replace the title with 'Monopoly demand'. You're finished; the demand curve for the market is the demand curve facing the monopolist, as there are no other providers of the good. The main difference compared with a PCM firm, as shown in the chapter on marginal revenue, is the marginal revenue curve for the monopolist firm. Figure 28.1 returns to the demand, average and marginal revenue curves for a monopoly firm.
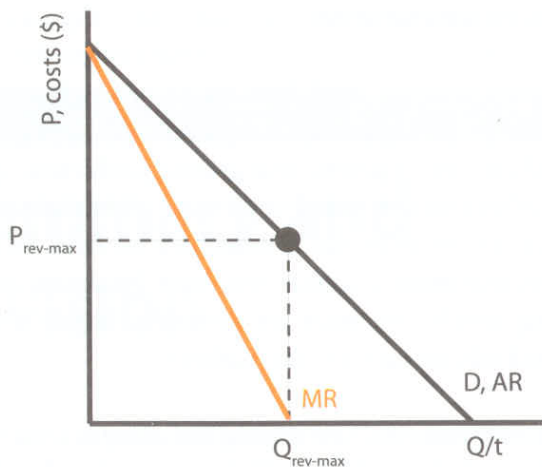
Figure 28.1 Demand, AR and MR for a monopoly

The market demand is the same as the monopolist firm's demand.

$AR = \frac{(P \times Q)}{Q}$, which means that AR is equal to demand.

$MR = \frac{\Delta TR}{\Delta Q}$. As the single-price monopoly will lose revenue as well as gain revenue by lowering price, the MR curve will fall twice as fast as AR.

TR is maximised at output level where MR = 0 which is where PED = 1.

Profit maximising price and output

As the monopoly firm faces a market where no other goods are substitutes, the market demand will be satisfied by the monopoly firm. And since we retain the assumption of the firm being a short run profit maximiser, the firm will produce where MC equals MR. By drawing the marginal cost curve of the monopoly firm we get the profit-maximising level of output seen in Figure 28.2. The single-price monopolist sets output at the profitmax point (MC = MR) which is shown by $Q_{\Pi\text{-max}}$ in the diagram. The price can of course be set anywhere below the demand curve at the profitmax point of output, but the monopoly will set the maximum price – denoted by the boundary of demand at $Q_{\Pi\text{-max}}$, which is at $P_{\Pi\text{-max}}$.
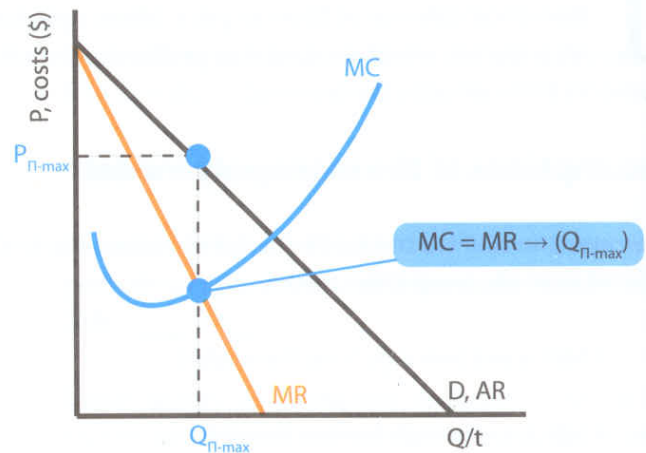
Figure 28.2 Profit-maximising level of output in a monopoly

The monopoly operates under the same set of criteria as the PCM firm; output is set at the level where **MC = MR**; $Q_{\Pi\text{-max}}$.

However, since the monopoly has price-setting power, the price is set above **MC** but within the boundary of the demand curve. Price is set at the highest possible point on the demand curve; $P_{\Pi\text{-max}}$.

Conclusion: P > MC.

The unit-cost picture and the market picture meld (= join) in the case of monopolies. The demand curve is a 'given' but since there is only one supplier, the assumption of profitmax and the entailing MC = MR condition turns the monopoly firm's MC curve into the market supply curve. I actually recommend avoiding the use of 'supply curve' in conjunction with monopoly output.

DANGER MINES **WARNING !**

The two most common mistakes in dealing with monopoly diagrams are 1) drawing faulty MR curves, and 2) confusing revenue maximum with the point where the MC curve intersects with the demand curve.

Faulty MR curves: The **diagram on the left** below illustrates the first problem. As the MR *must* lie 'halfway' between the demand curve and the price-axis, **all the MR curves in blue are incorrect**! Only **the orange MR curve is correct**. It is amazing how many times one sees this mistake in both economic literature and official exam papers. It's the sort of mistake that, when done by a student, is taken to show sloppiness, lack of understanding, or both. Like drugs, just don't do it.
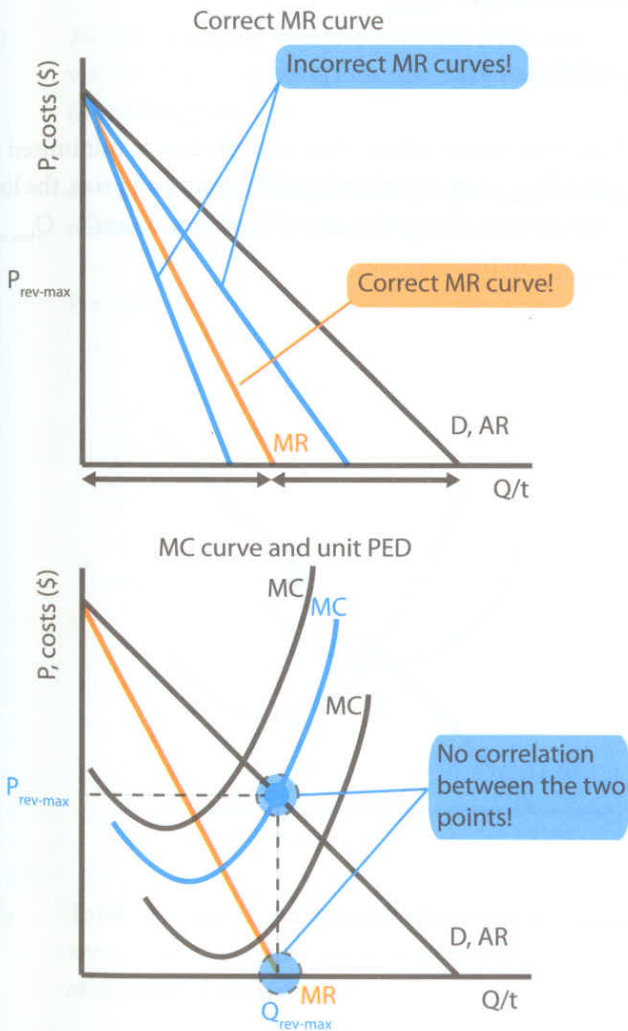
Figure 28.3 Good diagrams to illustrate monopolies

Revenue maximising price and output

While the profit maximising decision for a monopolist is not different from that of the competitive firm, the pricing decision leads to a higher price than marginal cost, as explained above. In many – but far from all – cases, the monopoly firm will have an average cost level that is below the price, as shown in Figure 28.4 below, giving the firm an abnormal profit (grey area).
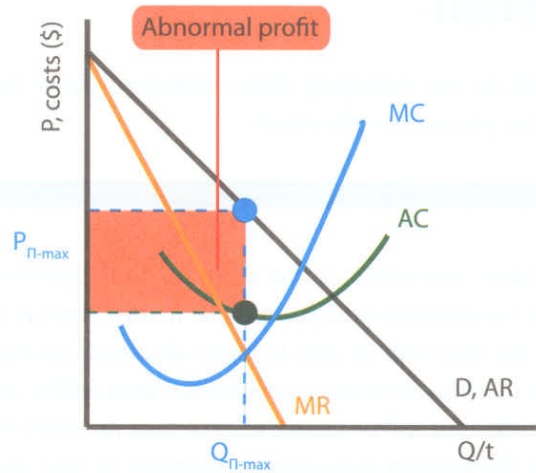
Figure 28.4a Abnormal profits and LR in monopoly - Profit-maximising monopoly

Output: Q is set at $Q_{\pi\text{-max}}$. This is the level where MC = MR.

Price: The monopoly will set the highest possible price allowed by market demand – $P_{\pi\text{-max}}$.

Profit: Entry barriers and lack of substitutes enable the monopoly to reap abnormal profits in the long run. This is shown by the blue area in the diagram; $P_{\pi\text{-max}} \times$ (AR – AC).
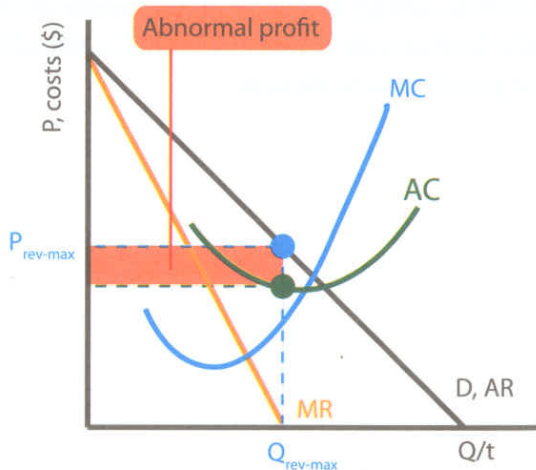
Figure 28.4b Abnormal profits and LR in monopoly - Revenue-maximising monopoly

The monopoly can forego short run profits in order to maximise revenue.

Revenue is maximised when output is set at the unit elastic point on the demand curve – which will also be where MR = 0. Price decreases and output increases compared to profitmax price and output. **Revenue increases and profit decreases compared with profitmax price and output.**

Question:

How well do the monopoly firms illustrated above fulfil the criteria for productive efficiency?

Figure 28.5 Normal profit and loss in monopoly

Since a basic assumption is that there are no competitors – i.e. there are no substitute goods – and that market entry is severely limited, the firm will be able to enjoy abnormal profits in the long run in addition to short run profits. Firms will be attracted to the market but will be unable to overcome the entry barriers, allowing the existing monopoly to continue to reap abnormal profits.

In Figure 28.5: Output is set according to profitmax/loss-min criteria, resulting in $Q_{\pi\text{-max}}$.

AC_0: The firm makes a normal profit.

AC_1: The firm makes a loss. Note that the loss is minimised at $Q_{\pi\text{-max}}$ and $P_{\Pi\text{-max}}$ – at any other level of output and price, the loss would be greater. Hence the use of "loss minimum" – $Q_{\text{loss-min}}$ and $P_{\text{loss-min}}$ – in the diagram.

Normal profits and loss in a monopoly

It is possible that the monopoly firm faces market demand that yields less than abnormal profits. Figure 28.5 shows two possibilities. In the case where the AC curve is tangential to the demand curve (AC_0) the monopoly will make a normal profit, as AR equals AC. If average costs are pictured by AC_1, then the total cost per unit is higher than total revenue per unit; the firm will run at a total loss represented by the grey area in the diagram. In keeping with the truism of profitmax and loss-min, the firm with an average cost curve of AC_1 cannot set the price at any other level in order to lower losses; MC = MR will render the lowest possible loss to the firm.

Summary & revision

1. The **assumptions** in the monopoly market are:

    1. One firm controls the market;

    2. There are very high barriers to entry;

    3. The firm is a short run profit maximiser

2. **Entry barriers** include the possibility of technical barriers (proprietary production methods), control of vital factor inputs (raw material such as bauxite for aluminium), legal barriers (intellectual property rights or a state monopoly on tobacco), huge economies of scale (natural monopolies) and the possibility of predatory pricing and/or cartels.

3. As the monopoly enjoys protection from market entry via barriers to entry the firm can make **abnormal profits in the long run**.

4. In a monopoly, **average revenue equals price** – thus, the demand curve is also the average revenue curve.

$$P = AR = D$$
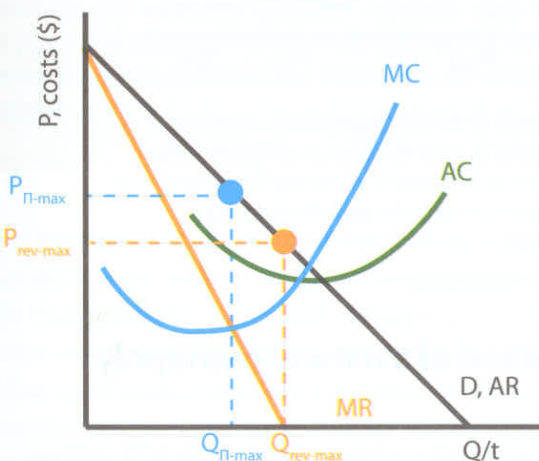
5. **Total revenue is maximised** where average revenue is zero – this is the same quantity where the price elasticity of demand is unitary.

6. **Profit maximising level of output** is where marginal costs equal marginal revenue. $MC = MR$; $Q_{\Pi\text{-max}}$. Thus, any profit maximising price will always be set along the elastic portion of the demand curve.

7. Quantity at revenue maximising output will always be greater than the profit maximising level of output. $Q_{revmax} > Q_{\Pi\text{-max}}$.

29. Natural Monopoly

Key concepts:

- Economies of scale
- Definition of a natural monopoly
- Different possible pricing and output decisions

Economies of scale

When there are very large fixed/initial costs and a very large potential for economies of scale based on technological gains, a monopoly might well have a profitmax point which renders a lower price and higher output than a perfectly competitive market. We now relax our earlier two assumptions added-on in comparing a monopoly with a perfectly competitive market. The sum of the MC curves for PCM firms are *not* the MC curve for the monopoly; and there are indeed economies of scale to be had.
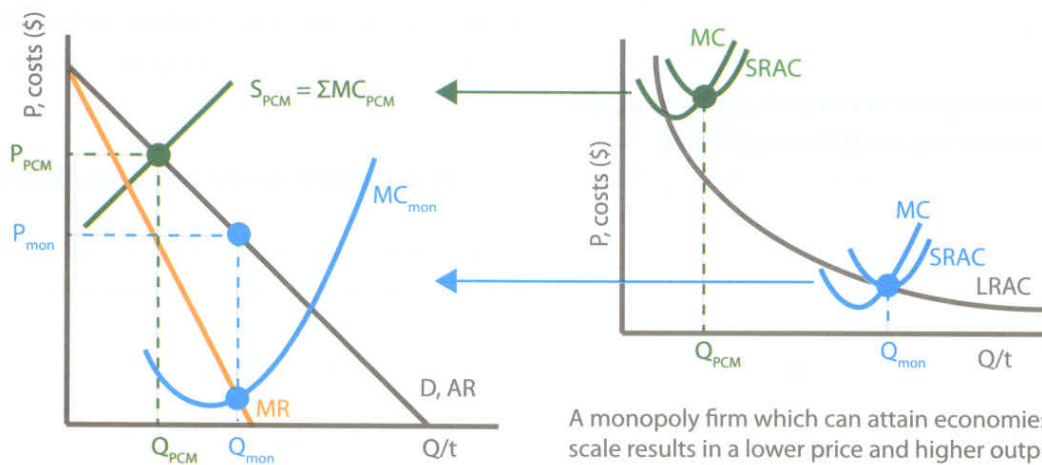
A monopoly firm which can attain economies of scale results in a lower price and higher output than is possible when competing firms have too small a portion each to benefit from scale economies.

Figure 29.1 Economies of scale in a monopoly

Definition of a natural monopoly

Figure 29.1 illustrates the outcome when a monopoly can enjoy economies of scale. The supply curve for the perfectly competitive market (left diagram) is the sum of the MC curves shown in the LRAC diagram to the right. Scale economies allow the monopoly to move along the LRAC (diagram on the right), allowing the monopoly to decrease MC and AC. The monopoly can produce a higher quantity and lower price – Q_{mon} and P_{mon} – than the perfectly competitive market. Simply put, a large quantity of (smaller) competing firms could not attain large enough scale to be able to compete with a single monopoly firm.

A natural monopoly has nothing to do with 'organic', in case any of you neo-hippies were wondering. A natural monopoly is when the barriers to other firms entering the market are in some way built-in to the environment, infrastructure or the nature of the good itself. In this, the single firm will have very large **benefits of scale** which competing firms could never acquire – this enables the incumbent (= "sitting", "current") monopoly firm to continuously undercut potential rivals and (intentionally or unintentionally) prohibit them from entering the market.

The concept of a natural monopoly is an extension of the scale benefit model covered above. However, a natural monopoly will have such large scale benefits, that no single firm can fully exploit them. 'Natural' in usage here means that the economies of scale make it more efficient for one firm to produce the good than for any other constellation in the industry.[1] Reasonably obvious examples of natural monopolies are gas networks for households, water lines, electricity, telecommunications and cable TV – imagine the costs in having each individual gas or electricity company having their own lines into each house! My favourite example is railways; imagine having two competing companies with their own tracks – side by side – running between two cities. How about two bridges…side by side?!

There are of course very high barriers to entry in industries characterised by natural monopolies. This is due to two common characteristics of natural monopolies:

1. Extremely high fixed costs and low variable costs

2. High initial (start-up) costs – often infrastructural

> **Definition: 'Natural monopoly'**
>
> A **natural monopoly** is a situation (not a firm!) where the market is most efficiently served by a single firm since the benefits of scale are so large that not even a single firm can fully exploit the potential economies of scale. Gas, water and electricity lines to homes are examples of natural monopolies.

On one of my trips through Ireland in the mid-eighties, I visited a far-northern town called Glencolombcille. It soon became apparent that getting there involved taking a series of buses – which were mostly provided by non-competing bus companies which had local monopolies for a number of town-to-town routes. It also became apparent that there was no cooperation between the companies since arriving buses invariably arrived ten minutes after connecting buses had departed![2] Bus services

have high fixed costs since it basically costs the same to drive an empty bus from point A to point B as a bus full of inebriated Irish fiddle-players and confused tourists. The route I was travelling simply could not have supported competing bus lines, and it would be ludicrous to think that competitive forces could have lowered the market price while firms still covered their costs.

The *Indian Pacific* at the Forrest 'passing place' in the middle of the Nullarbor Plain, Australia - the longest stretch of straight railway track in the world. The train runs from Perth on the shores of the Indian Ocean to Sydney on the Pacific Coast. The trip takes 3 nights and 4 days and covers 4,352km - there are no competing trains on this route!

Natural monopolies do tend to occur where fixed costs are very high, giving potential economies of scale only to firms which can exploit sheer size to some extent and spread the fixed costs over large output; one firm can satisfy market demand at a lower cost than competing firms. The markets for cable TV, water/ gas/ electrical utilities, railroad tracks and local bus services all have strong elements of natural monopoly. Just imagine how inefficient it would be to have several hundred competing gas pipes to each household or fifteen firms each laying down competing railroad tracks! The Öresund Bridge connecting Sweden to Denmark is also a natural ~~disaster~~ monopoly[3] … and *two* bridges would basically be a monument to financial stupidity – and that might explain why politicians at this very moment are planning yet another connection via a tunnel[4].

---

1 It is noteworthy that the term natural monopoly does not in fact refer to the dominant *firm* but rather to the *industry* itself. There can be many competing firms in a natural monopoly situation and in using the term natural monopoly we are claiming that one firm would increase efficiency in the industry. This is a common political argument put forward in favour of nationalising water works, electricity and train services.

2 I strongly suspect collusion between the bus companies and local pubs – which were always right next to the bus stations. It was all very Irish and very fun; we'd all get off one bus, tumble

into the pub for a few pints, singing, joke-telling and fiddle-playing, and off we'd go on the next bus. I spent more in the pubs than on the bus tickets.

3 The bridge was initially a monumental failure in terms of recouping the immense initial costs of building it. It ran at a loss for almost seven years since its inauguration in July 2000 – in part because of the high price for those in the vehicles that use it.

4 I am not making this up!

Different possible pricing and output decisions

The issue of is illustrated by the continuously falling LRAC curve in Figure 29.2 as average costs fall throughout the entire boundary of market demand, LRMC too continues to fall. This renders three possible pricing outcomes – all of which are lower than for a PCM:

1. **Profitmax pricing;** where the monopoly sets price according to the MC = MR condition. This creates an output of $Q_{\Pi\text{-max}}$ and a price of $P_{\Pi\text{-max}}$. The firm makes an abnormal profit of the blue area. Examples include private telecoms companies and gas companies.

2. **Average cost pricing;** is when the monopoly sets the price at the break-even point of AC = AR.

3. **Marginal cost pricing;** at an output of Q_{MC}, the monopolist would incur a loss of the purple shaded area. Marginal cost pricing is commonly done in order to achieve social benefits – for example, providing water – and often involve public ownership or government funding via subsidies.

If the industry is forcibly turned into a duopoly, each firm would have AC_{duo} and losses (grey area).

Profitmax: the monopoly enjoys an abnormal **profit** (blue area) at $P_{\Pi\text{-max}}$ and $Q_{\Pi\text{-max}}$.

AC pricing: The firm sets output at **break-even** point, earning a normal profit.

MC pricing: Setting price at MC means a **loss** (purple area) for the firm, which is only tenable in SR. (Or via subsidies and price discrimination in LR)

Fig. 29.2 Natural monopoly

What would happen if the natural monopoly was broken up in order to establish a competitive market? Assume an existing monopoly at the profitmax price of $P_{\Pi\text{-max}}$. If government forced the firm to break up into two parts – a **duopoly** – each firm would have a far higher AC than the monopoly; the two firms would have an average cost of AC_{duo} and an output of Q_{duo}. This is clearly economically impractical. (Note the suboptimal *productive efficiency* in all three pricing scenarios, and the suboptimal *allocative efficiency* in all cases but MC pricing.)

Public and merit goods

Goods provided and/or produced by government monies, i.e. **public and merit goods** such as roads and health care are often priced at AC or MC. Societal welfare is deemed to benefit so much by the provision of such goods that the profitmax condition is not considered applicable. Instead, the good is priced at cost or – via subsidies – at below marginal cost.

Whether public/merit goods warrant government support or provision or not is subject to very heated debate. The trend over the past 20 years has been increasingly in favour of market openness and privatization, where a number of traditionally government-controlled markets have been sold off or otherwise

subjected to open market competition. State-owned telecom companies, electricity and railroads are notable examples. In certain cases the results have been rather dismaying, for example England's privatisation of British Rail, completed in 1997, was called a 'Disaster' by *Time Magazine*[5] ; over 100 different competing rail-service providers saw costs spiralling upwards continuously leading to fewer departures and lower levels of service. There are increasingly loud cries for re-nationalisation. A very similar debate raged in California during 2002, as overloaded private electricity companies led to daily blackouts. You can just imagine how popular sudden power interruptions were to the programmers in Silicon Valley.

Summary & revision

1. A large monopoly might enjoy significant **economies of scale** and therefore be able to produce at a lower price and higher quantity than competitive market firms.

2. A **natural monopoly** is a market *situation* (rather than a specific firm) where there are considerable entry barriers (very high fixed and start-up costs) and no single firm would be able to fully exploit these scale benefits. Common examples are electricity and gas lines.

3. The economies of scale result in **falling or constant average and marginal costs** throughout the realm of demand.

4. A natural monopoly can set price and output at different levels: 1) At **profitmax** price and output (MC = MR); 2) At **breakeven price** and output (AR = AC); 3) At a level considered approaching the **societally optimal** – where a loss can be made up for via subsidies or price discrimination (AR = MC).

5. Natural monopolies are **frequently controlled/operated by government**. Price and output is commonly set at breakeven or marginal cost price to maximise social benefits. Good examples of government run natural monopolies include railroads and electricity networks.

---

5 *Time*, July 14 2003, pages 54 – 57.

30. Monopoly vs. Perfect Competition

Key concepts:

- Disadvantages of monopolies
- Efficiency in monopolies and perfectly competitive market firms
- And yet...
- Regulating monopoly power

It's fairly safe to say that monopolies have a rather poor reputation amongst both us consumers and also the political entities we elect to safeguard our interests. But the full story, as so often is the case, shows that monopolies in fact have a number of advantages – sometimes outweighing the disadvantages. The ability to compare different market structures using economic jargon and diagrammatical illustrations is the MOST important component of theory of the firm. Simply committing a number of diagrams to memory will not enable you to address the analytical questions posed in the IB exams – you must be able to utilise the concepts and diagrams to explain, assess and analyse the world around you. Study this section carefully.

Disadvantages of monopolies

In comparing the perfectly competitive firm/market with the monopolistic firm, we must make a few additional assumptions. This is not done in order to 'make the diagrams work' as some of my 'snottier' students claim, but to enable us to compare apples with apples, rather than apples with oranges. We make two assumptions in addition to those previously outlined:

1. The sum of the perfectly competitive market's MC curves are identical to the supply-curve on the perfectly competitive market and also to the **monopoly's MC curve; $\Sigma MC_{PCM\ firms} = S_{PCM} = MC_{monopoly}$**. This is not as outlandish as first sight might indicate. Consider a competitive market where all firms are efficient and operating along the best possible MC curves and that one large firm buys up all the individual firms. The large monopoly runs each separate firm as a plant belonging to the mother company – the MC remains the same, since we also assume that...

2. ...there are **no benefits of scale** involved. If there were, we could not assume that $\Sigma MC_{PCM\ firms} = MC_{monopoly}$, as the monopoly would gain scale benefits and thereby lower both AC and MC.

The Lockheed Constellation, part of the duel for monopoly of air routes in the post war period

Higher price and lower output

It follows common sense that a firm which can affect supply but not demand will set a price which will optimise profit by way of increasing the margin (= distance) between the market price and the firm's costs. This is done by restricting output and setting the price above marginal cost. Figure 30.1 shows that a monopoly will have a different outcome in terms of price and output. The MR curve for the monopolist will be lower than demand at all levels which sets the MC = MR point at lower output than would be the case in a competitive market. Note that the monopoly MC curve is the sum total of all PCM firms' MC curves, which means that the supply curve for both the competitive market and the MC curve for the monopoly are one and the same.
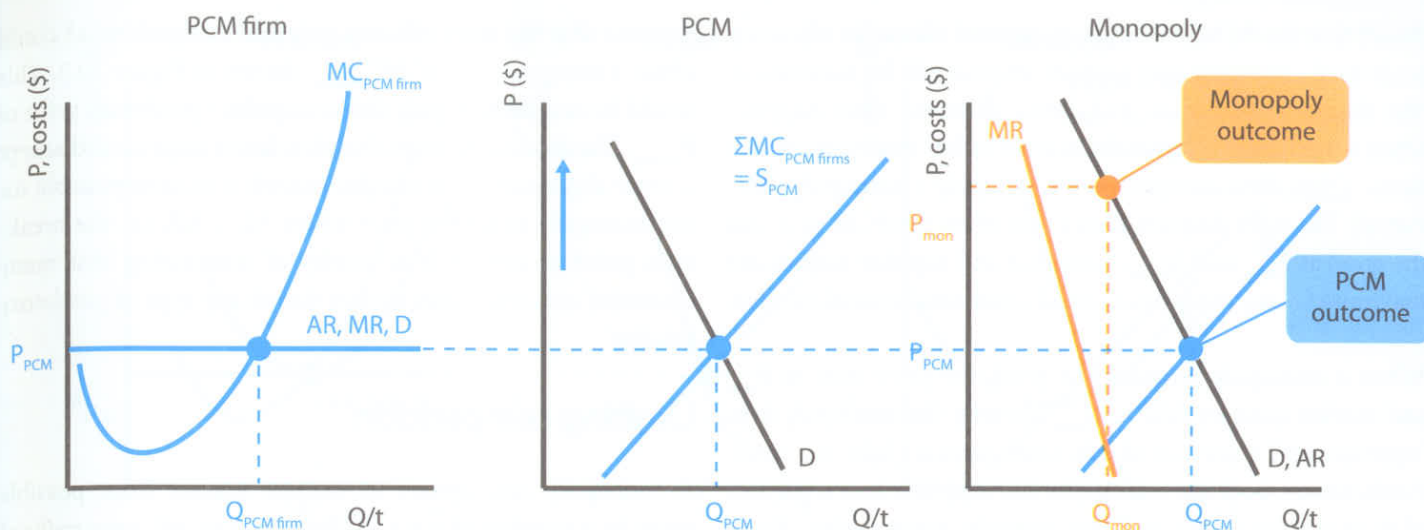
Fig. 30.1 The PCM firm compared to a monopoly

As the monopoly restricts market output and raises the price, the producer gains at the expense of the consumers. This loss of *consumer sovereignty* is the most noticeable negative effect of monopoly market situations; consumers 'pay more for less'. See below for more depth in *Outside the box: Monopolies and deadweight loss*.

OUTSIDE THE BOX

Monopolies and Deadweight Loss

I do apologise if the additional theory in these boxes weigh you down by seemingly adding to your already considerable workload. My justification is that I am trying to provide you with a set of tools which will enable you to better address the questions you will face in the field of economics. Being able to use the concept of consumer/supplier surplus to show deadweight loss is quite valuable.
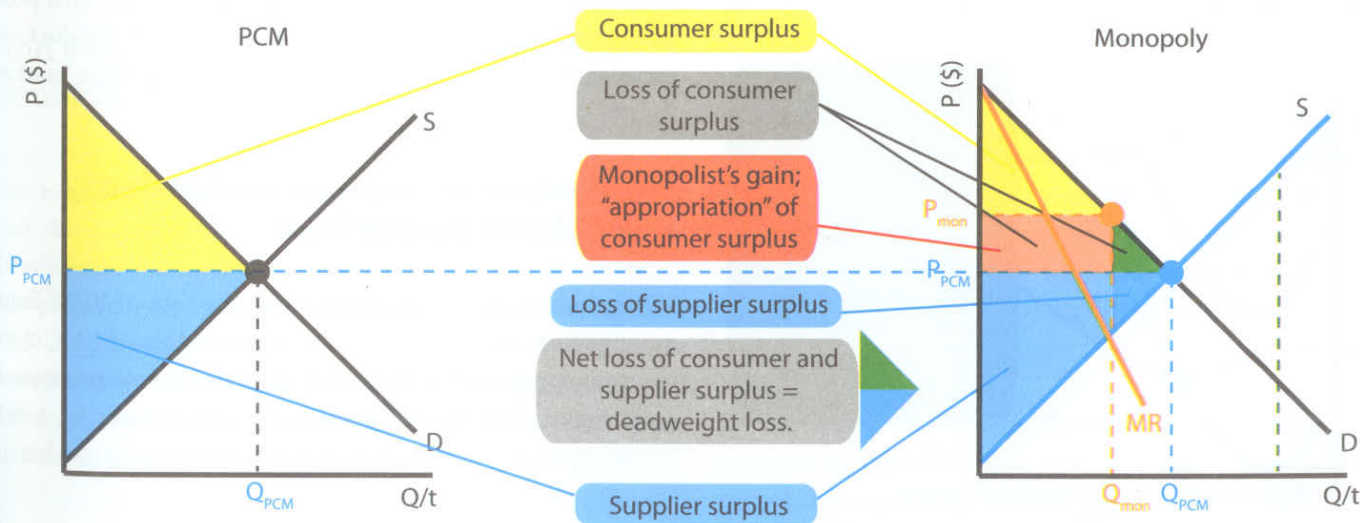
Fig. 30.2 Deadweight loss in a monopoly

Recall that the PCM will result in optimal allocative efficiency since total consumer and supplier surplus will be maximised. The diagrams in Figure 30.2 below illustrate what happens when a monopoly supplies the market rather than competitive firms. (Note the assumption of PCM supply = monopoly's MC curve!) The right diagram shows that when the PCM provides the good at P_{PCM} and Q_{PCM}, consumer and supplier surplus are maximised – any other price would lower total societal surplus.

When a monopoly provides the good, the price rises to P_{mon} and market quantity falls to Q_{mon} whereby the monopoly firm "appropriates" a portion of consumer surplus (light red area). Furthermore, both the consumers and suppliers lose a portion of their surplus (green and blue triangles respectively). These triangles show that at each level of output lost, the potential benefits (shown by the D-curve) outweigh any increase in cost (shown by the MC curve). This is a welfare loss to society commonly termed a **deadweight loss**.

The PCM has no deadweight loss and is therefore allocatively efficient. The monopoly's deadweight loss illustrates that it is not allocatively efficient.

Predatory pricing

One way to create and/or uphold a monopoly is simply to get rid of rivals and dissuade (= discourage) potential entrants to the market. A monopoly has the market power to price at a lower level than potential entrants can set. This pricing policy is short run, as it is intended to kill off rivals (hence the term "*predatory pricing*") so that the monopoly can maintain profitmax pricing and abnormal profits.

The monopoly firm can set the price as low as P_{min}. This is the break-even point, i.e. where AC = AR. The monopoly would make a **normal profit**. Potential entrants cannot match this price, which dissuades market entry.
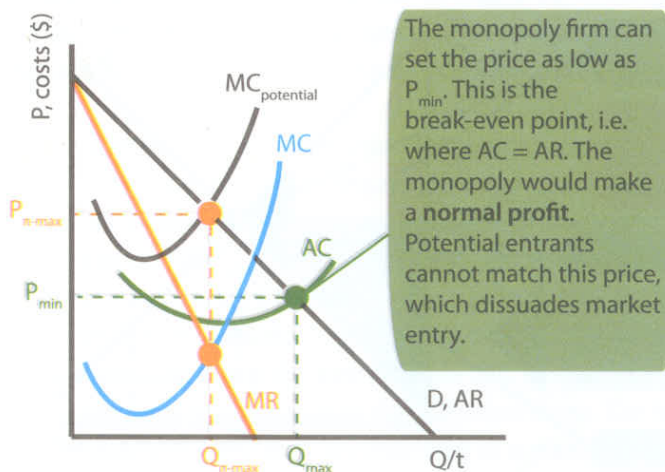
Figure. 30.3 Predatory pricing in monopoly

Suppose that the most efficient *potential* competitor(-s) could attain a marginal cost of $MC_{potential}$, shown in Figure 30.3. This would be competitive with the monopolist's profitmax price of $P_{\Pi\text{-max}}$. The monopoly might therefore lower its price and accept a lower abnormal profit margin. Indeed, it is quite possible for the monopoly to set the price where AC = AR, i.e. the break-even point of output. It is worthy of mentioning that many countries' anti-trust laws strictly forbid any type of predatory pricing.

Limiting competition

A monopoly can protect its captive market from possible entry in a number of ways. It can refuse to sell monopolised parts to potential competitors, which is what Microsoft did in the late 1990s when the company refused to sell Windows to computer manufacturers which did not also accept the Web browser Explorer, which ultimately cost the leading browser at the time, Netscape, most of its market share. In essence, a monopoly can seek to block/hinder rivals' access to supply-chains in production and retail/outlet chains at the sales end of operations. Monopolies also wield considerable financial and political power, which can create a 'cosy' relationship with governments which not only are consumers of monopoly goods, but are also responsible for regulating monopolies. Monopolies are also interested in setting the end-user price (i.e. the retail price) and can therefore attempt to force retailers to set prices which are non-competitive.

Higher costs and less innovation

It is possible that a monopoly firm may choose 'the quiet life'[1], i.e. be satisfied with a comfortable level of abnormal profit rather than putting effort into increasing efficiency. This profit saticficing outcome would be suboptimal, since productivity increases would be less than in a perfectly competitive market.

Efficiency in monopolies and perfectly competitive market firms

We again include the assumption that the sum of PCM firms' marginal costs equals PCM market supply and also the MC curve of a monopoly, e.g. the monopoly does not have economies of scale. Figure 30.4 below compares the cost pictures for a firm operating on a competitive market and a monopoly market.

---

1 Description of monopoly by Nobel laureate J. R. Hicks. web. mala.bc.ca/econ/quotes.htm
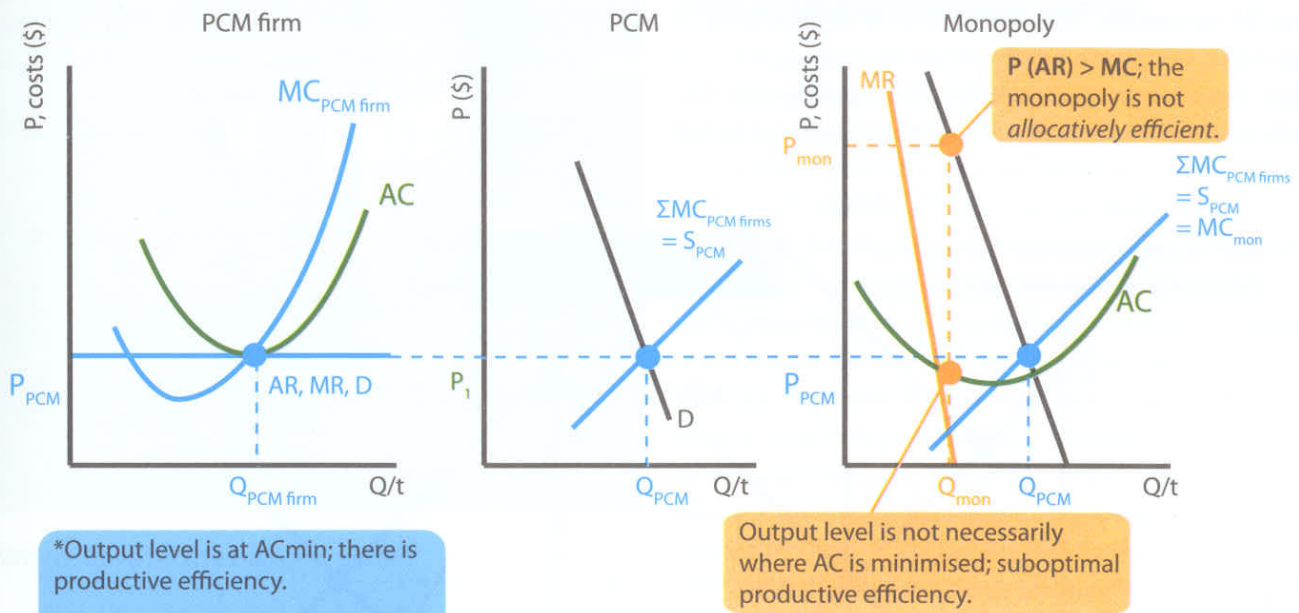
*Output level is at ACmin; there is productive efficiency.

*P (AR) = MC; there is allocative efficiency.

Output level is not necessarily where AC is minimised; suboptimal productive efficiency.

P (AR) > MC; the monopoly is not *allocatively efficient.*

Figure 30.4 Efficiency in monopoly and the PCM

- **PCM firm:** Recall that the long run equilibrium of the competitive market is when each PCM firm is operating at the lowest possible average cost (AC_{min}) and where the price of the last unit is equal to the cost of the last unit (P [MR] = MC). The PCM firm is therefore both *productively* and *allocatively efficient*.
- **Monopoly:** The monopoly firm is in stark contrast, since the price is higher than marginal cost; the monopoly is *not allocatively efficient*. Nor will the monopoly firm necessarily be productively efficient, as the output level which has been determined by the MC = MR condition means that the monopolist might well produce where AC_{min} is not attained.

And yet... a closer look at some possible benefits of monopolies

Life is seldom so simple as to neatly divide reality into good or bad[2]. Monopoly power is no exception as was shown in Chapter 29 where the monopoly can prove itself superior to the perfectly competitive firm: 1) economies of scale; 2) natural monopolies; and 3) government run monopolies. Additional possible benefits of monopolies are increased profits for research and development (R&D), the reduction of negative externalities and the process of "creative destruction".

Profits for R&D (innovation and new products)

There is also the possibility of monopolies having far greater research and development (R&D) capabilities as a direct result of being able to earn abnormal profits. Ploughing profits back into product enhancements and innovative new products could increase consumer choice and benefits to society in the longer term. PCM firms could never have the LR profits available to monopolies, and would thus not be able to fund R&D to the same extent as monopoly firms.

Lower negative externalities

Finally there is the distinct possibility that the monopolist outcome of higher prices and lower output actually benefits society! This would be the case in the production and consumption of goods which are harmful to society, called **negative externalities**. For example, the production of certain plastics is highly harmful to the environment, as numerous carcinogenic toxins (= poisons) are released into the air. A monopoly firm might well produce less of the good and thereby have less negative impact on society than would a perfectly competitive market.

As previously shown it is theoretically likely that a profit maximising monopoly firm will set a higher price and lower output than a perfectly competitive firm. This would normally result in sub-optimal allocative efficiency (and a dead-weight loss) since the price (AR) would be higher than MC. (See Figure 30.1, *The PCM firm compared to a monopoly*.)

2   That is why we have divorce and divorce attorneys.

However, let us assume that the monopoly is producing a good for which the negative externalities are considerable, for example a major tobacco firm wielding monopoly power. The good inflicts costs to society in increased sick days and falling productivity, increasing health costs etc. This is shown in the left figure in figure 2.4.6 as the negative externality; MPC *δ* MSC.
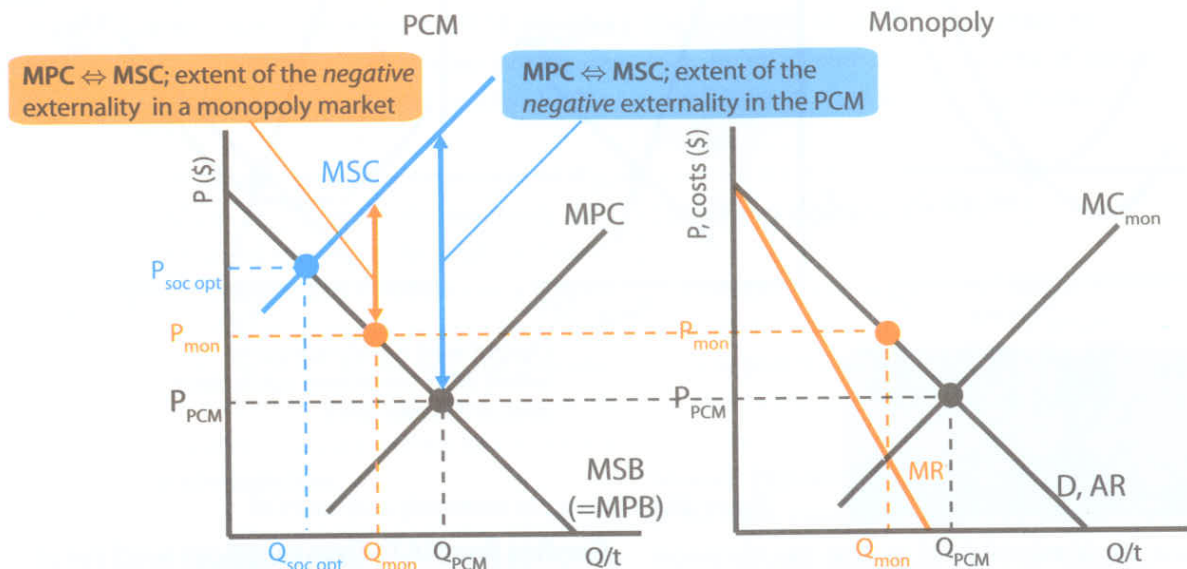
Figure 30.5 Lower negative externalities due to monopoly

Assuming (as was done in comparing the PCM with a monopoly) that the sum of PCM firms' individual cost curves (**MPC** in the left diagram) would be the same as the MC curve for the monopoly (**MC$_{mon}$**), then the **MPC** is the same as the **MC$_{mon}$**. This would render a competitive market output of **Q$_{PCM}$** and a price of **P$_{PCM}$**, and a *negative externality* (double arrow in diagram).

The monopoly plastic company, on the other hand, would base the output/price decision on a different MC = MR point. The right diagram shows the **MC = MR** point where the monopolist would set output, and the higher monopoly price of **P$_{mon}$** and a lower output level of **Q$_{mon}$**. Since this results in a lower output and higher price (**Q$_{soc\,opt}$** and **P$_{opt}$**) for the monopoly, the negative externalities of producing plastic in the monopoly are actually lower – shown by the beige double arrow. In other words, the monopoly is actually more allocatively efficient than the competitive market.

I mean, let's face it; a goodly number of countries' governments create government run monopolies for tobacco and alcohol (Sweden, Norway and Finland for alcohol and Greece and Spain for tobacco) for a good reason – other than tax revenue that is. Operating under monopoly conditions allows

governments to price these 'bads' more in keeping with societal costs and benefits. When my home country Sweden instated a government monopoly on the sales of beer in 1976 alcoholism, alcohol induced psychoses and intoxication fell by some 40% amongst teenagers.[3]

Other possible advantages of monopoly

A rather interesing possibility was put forward by noted economist Joseph Schumpeter, who theorised that there is an on-going process of **creative destruction**. While it sounds distastefully 'socio-Darwinist', Schumpeter posited that over time sheer monopoly power would force potential entrants to develop outstandingly new/superior substitute goods in order to compete in some way, rather than the incremental (= 'little at a time') improvements of competitive markets. The new and major new product would effectively destroy the market for the 'old' product while creating a new market. These 'leaps' of innovation and invention effectively terminate markets – creative destruction – in providing superior substitutes. This is posited as being more beneficial in the long run than competitive markets.

---

3   See Mats Ramstedt's article at www.nordicwelfare.org

"Where have all the buggies gone?"

Monopoly versus PCM – a summary

It has been shown that, under the assumption that the PCM supply curve is identical to the monopolist's MC curve, the monopoly will:

- enjoy *abnormal profits* in both short and long runs
- be *allocatively inefficient* (P [AR] > MC), and in all likelihood *productively inefficient* ($Q \neq AC_{min}$)
- produce a lower level of output at a higher price – this leads to a net loss of consumer and supplier surplus called *deadweight loss* [not part of syllabus – see 'Outside the box; Monopolies and deadweight loss']

In relaxing the assumption of $S_{PCM} = MC_{mon}$, there is a mixed bag of both disadvantages and advantages. A monopoly could:

- raise existing *entry barriers* and indulge in *predatory pricing* to dissuade entrants
- idle along at low productivity levels, resulting in *sub-optimality* in productivity gains
- enjoy *economies of scale* and *natural monopoly*, lowering price and increasing output compared to the PCM
- *lower the negative externalities* of production and consumption of goods
- be *more innovative* that competitive firms, since abnormal profits are available for *R&D*

Regulating monopoly power

Most countries will have legislation which severely limits both the creation and activities of monopolies and other restrictive competition practices. As we shall see in the chapter on oligopolies, when the few act as one there is a monopoly, i.e. when oligopolistic firms act in concert – by perhaps forming a cartel – the result is basically a monopoly. These anti-trust laws are in place to limit and regulate such collusive (= cooperative) and harmful behaviour. The anti-trust laws in America are amongst the harshest in the world, and make it illegal for a firm to have monopoly power at all (!), whereby firms that are judged to be monopolies can be forcibly broken into parts by order of law. The most famous example perhaps is when Standard Oil, created by John D. Rockefeller, was forced in 1911 to divest (= sell) itself of holdings in many other companies – one of which became today's ExxonMobile. This is also what happened to the aforementioned ALCOA in 1945 and AT&T in 1983; both were considered to have too large a market share and were forced to break up into smaller competing firms. As a comparison, the European Union's charter does not explicitly forbid monopoly creation per se, but the 'abuse of the dominant position of a company which negatively affects the trade between Member States'[4].

POP QUIZ 30.1

Monopoly

1. Explain why a monopoly can earn a supernormal profit even in the LR.

2. Explain why the profit maximising monopolist will always set output along the elastic portion of the demand curve. Use a diagram to illustrate your answer.

3. Explain how/why a monopoly which aims to maximise revenue will have a higher level of output than when it is profit maximising.

4. If, at a certain output, the MC is greater than MR, what should a profit maximising monopolist do?

---

4 http://europa.eu.int/comm/competition/antitrust/legislation/entente3_en.html

5. Is it always the case that a monopoly leads to a higher market price and lower output than a perfectly competitive market or are there exceptions?

6. Explain how a natural monopoly's price would differ from that of competitive firms on the same market.

Summary & revision

1. **Assuming there are no economies of scale** available for a monopoly, the profit maximising monopoly firm will set lower quantity and higher price than the perfectly competitive market firm.

2. The higher price and lower output level of the monopoly result in a net loss of both consumer and supplier surplus (e.g. societal surplus). This **deadweight loss** illustrates that the monopoly is allocatively sub-optimal.

3. A monopoly can also engage in so-called **predatory pricing** where price is set at breakeven (AR = AC) and potential rivals/entrants are unable to compete at this price.

4. Monopolies can **uphold a monopoly situation** by limiting access to factor inputs such as specialised components and blocking supply/retail chain networks.

5. It is theoretically possible that monopolies choose to lead a '**quiet life**' and target acceptable levels of profit rather than strive for lower costs via increased efficiency.

6. Assuming there are no economies of scale involved for the monopoly firm in comparison with a perfectly competitive market, the profit maximising monopoly firm:

   a. Will set price above marginal cost (P > MC). The monopoly will be operating under **suboptimal allocative efficiency**.

   b. Will be **productively inefficient** since output will not be set where average costs are minimised (Q $\neq AC_{min}$).

   c. Will be able to enjoy **long run abnormal profits** due to high barriers to entry.

7. There are several 'howevers' in terms of the negative outcomes in monopolies. Advantages over the perfectly competitive market include:

   a. **Economies of scale and natural monopolies** might well have a lower price and higher output than would be the case in a perfectly competitive market.

   b. **Abnormal profits** from monopolies can be ploughed (plowed) back into research and development – **R&D** – resulting in innovation and new products.

   c. **Lower negative externalities** in production since the monopoly firm has a higher price and lower output than would be the case on a competitive market.

   d. The theoretical possibility of **creative destruction** – a monopoly might force potential entrants/competitors to take giant innovative leaps forward to be able to compete with the monopoly. The significant improvement of the substituting good would destroy the monopoly.

8. Most countries will have regulations in place to limit monopolies in general or curtail monopoly power at the very least. **Anti-trust laws** commonly limit market share and thus the extent to which firms are allowed to merge – either voluntarily or via hostile takeovers. Other regulations include penalties for firms engaging in predatory pricing, collusive behaviour and limiting access to raw materials for other firms.

31. Monopoloistic Competition

Key concepts:

- Assumptions of the monoplistically competitive market
- Degree of market power
- Profit in the short run and long run
- Non-price competition in monopolistic competition

> Microsoft has had clear competitors in the past. It's a good thing we have museums to document that.
>
> Bill Gates.

We have now covered the two extremes of perfect competition and monopoly. While many markets have strong elements of these market types, most goods will be provided by firms which have elements of both. The result is **imperfect competition**, since most goods are in fact not homogeneous and are sold using arguments which are not based on the price of the good. These forces of *non-price competition* provide the foundation for monopolistic competition, which is a mixture of elements from both the competitive and monopoly markets. I often draw a Venn diagram on the board – to groans of weariness from my poor kids – with one circle as a monopoly and the other as a competitive market.[1] They perk up a bit when I start doing the Gunnery sergeant Hartman routine; "This is my rifle! There are many like it but this one is mine!"[2]

The market structure in monopolistic competition can, once again, be 'tasted'; the term monopolistic entails the firm being alone in providing the product while competition implies that the firm still has to take into consideration the possibility of *substitutes*. Consider running shoes; there are many competing brand names such as Adidas, Brookes and Asics, but only one Airmax – produced by Nike. There are many hamburgers – but only one Big Mac; many taco-restaurants but only one Mama Rosita's; many wristwatches but only one Fortis B-42, and so on.

Assumptions of the monoplistically competitive market

The assumptions of a monopolistically competitive market are almost identical to those outlined in the model of perfect competition; many sellers; no entry or exit barriers; the firm is a short run profit maximiser; and the market actors have perfect knowledge. There is one important change, namely that the good is **non-homogeneous** or **differentiated**.

> **Definition: 'Monopolistic competition'**
> A market where numerous sellers provide goods which are non-homogeneous/differentiated is a **monopolistically competitive** market.

---

1 Heads hit the tables when I add on a third circle; oligopoly.
2 From the classic movie *Full Metal Jacket* by Stanley Kubrick

Degree of market power

The simple examples above together with the added assumption of differentiated goods outline the three main characteristics of monopolistic competition:

1. There are a **large number of providers** of the good on a most competitive market. While the PCM will have a large number of suppliers and the monopoly only one, the monopolistic market will have a range of suppliers depending on the good. The main point is that the firms are independent competitors with relatively small market share, therefore firms will largely disregard the actions of other firms and the incentive to collude is quite small.

2. **Products are differentiated** by way of marketing and advertising. While there might not be any tangible (= substantial) difference between a pair of Reeboks and a pair of Nikes, the marketing efforts behind these goods seek to convince the consumer that there is a real difference between the brands. Beauty is in the eye of the beholder, and by strengthening a brand name and image, the firm attempts to create a difference in the perceived benefits of its product. (See chapter on product differentiation in oligopolies, Chapter 33.)

3. The firm has a **certain degree of market power**. As the firm has the monopoly (or proprietary) rights to the product, or brand, a degree of brand loyalty can be created by careful profiling of the product/ brand image, which serves to separate the product from substitutes. The firm is therefore not a pure price-taker but a price-giver to a certain degree.

4. There are **no or low barriers to entry**. In other words, it is likely that successful firms will attract rivals with similar products.

It turns out that there are in fact few market constellations which live up to the above characteristics and create *pure* monopolistic competition. Most of the examples, upon closer examination, show a very high concentration ratio; the four to five largest firms on the market have a controlling share of the total market. This is in fact an **oligopoly** market, dealt with in the next chapter. To remain *strictly* within the definition of monopolistic competition, one often winds up with local goods and service providers. Pubs, restaurants, corner stores, car repair shops and hotels can be considered purely monopolistically competitive.[3]

---

3 My editor Rory, with whom I stayed for a week at his home in Australia organising this book, often picked up some very

230

Profit in the short run and long run

As the firm sells a differentiated product, it will face a downward sloping demand curve rather than a perfectly elastic one like the PCM firm. However, the element of *substitutability* will have strong impact on the price elasticity of demand for the firm. Recall that the closer the (perceived!) substitute, the more elastic demand tends to be. A defining element regarding short and long run for the monopolistically competitive firm is to what extent consumers regard a good as substitutable. Let's not get my editor sued for copyright infringement in our example and use the name 'mePhone' for our exciting new mobile device from the Orange Megacompany.

- **Innovative stage:** Assume that mePhone hits the market with a product which is indeed markedly innovative and has no close substitutes, giving the producer a monopoly market. This is shown in Figure 31.1, where our innovative firm mePhone enjoys an abnormal profit, owing to the fact that other firms have not yet entered the market with their own versions of the mobile phone.

- **Close-follower stage:** Ultimately, since we are assuming perfect knowledge/information and no entry barriers, other firms see the abnormal profits and start to produce copy-cat products which are close enough substitutes to steal market share away from the innovative firm. These 'close-followers', in providing substitutes, will cause a decrease in demand for the mePhone, shown by the shift from D, AR_0 to D, AR_1 in the next diagram. The price elasticity of demand increases since there are now substitutes available on the market. The mePhone firm has been forced to lower the price from P_0 to P_1, but has still lost customers to the degree where it is making a loss, shown by the grey area.

- **Long run:** In the long run, as other firms also accrue losses and leave, and potential entrants are dissuaded by losses or scant profits, the mePhone firm will see an increase in demand and decreasing PED. Over time, monopolistically competitive firms will produce where the AC curve is tangential to the AR curve, the result being that in the long run, the monopolistically competitive firm will make a normal profit only, just as the PCM firm.

---

yummy sandwiches from a local bakery. He spoke of how he always went to this bakery since the bread was a really tasty variety of wholegrain. Then we started talking about beer preferences...

mePhone – innovator stage
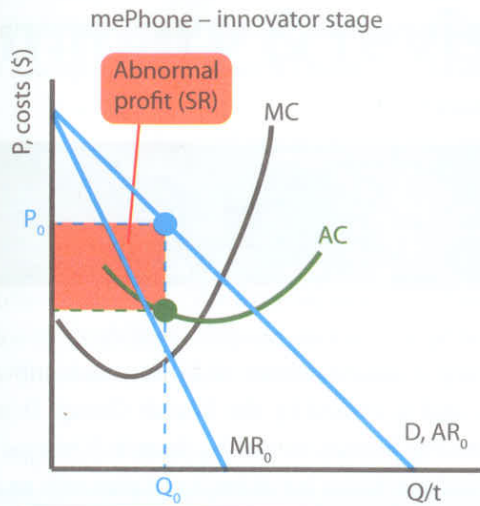
The innovative firm can make a SR abnormal profit, by virtue of being alone on the market.

mePhone – close follower stage
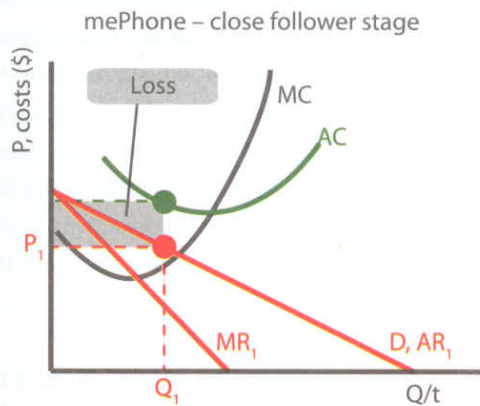
As other firms are attracted by the possibility of abnormal profit, the innovative firm's demand curve will shift to the left whilst substitutes increase the PED. The firm operates at an **intermediate** loss.

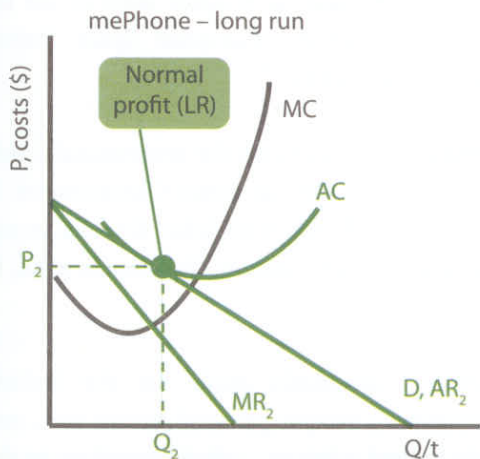## mePhone – long run

Dissuaded by possible or realised losses, firms leave the market, shifting demand for the innovative firm to the right and lowering PED as the amount of substitutes fall. **Normal profits** are made.

Figure 31.1 Monopolistic competition – SR to LR

Be careful in ascribing too large a degree of static equilibrium to the model! It is highly unlikely in the real world that the innovative firm will simply stand by and watch how its market share is eroded by competitors. The firm will strive to stay ahead by marketing its product and by continuously looking at ways to improve the good – superficially if nothing else. What the model shows, is that there will be a tendency towards the LR equilibrium point of zero abnormal profits.

Non-price competition in monopolistic competition

I have always been a sucker for new, cool consumer electronics. When Sony came out with the *Walkman*[4] in 1978/'79, I was first in line. It cost me $280 in 2003 prices and exchange rates, which is a bit pricey. Within a year of my buying it, numerous other electronics firms came out with similar products. Sony reacted in true monopolistically competitive style; by differentiating the product. Sony set about making many different styles and models, all equipped with new features different from other brands. Ten years later, Sony had over 200 models of Walkmans. The 1990s saw Sony extend the brand to portable CD-s (Discman) and then the Minidisc. MP3 players and iPods are the latest. Each product creates an array of 'me-too' products vying for a share of the market.

> **Definition: 'Product differentiation'**
>
> When firms on a monopolistically competitive market enhance product heterogeneity by changing the good (physically) and/or using advertising to create an image and uphold/enhance a brand, one speaks of **product differentiation**. This is a form of **non-price competition**.

Differentiating a product is seldom as hands-on and tangible as adding new gizmos and dials to the product. Often as not it is a matter of simply changing the packaging or colour. How many times have you seen a 'new' product advertised by some phrase like 'Now in an all-new Mango-Ice-Fire colour!' or other such nonsense?![5] In any case, the adding-on of tangible and/or intangible features allows the firm to pass its product off as differentiated in its marketing drive.

4   *Walkman* is an example of brand degradation, which is when a brand name becomes a generic term for any similar good. *Hoover*, *Aspirin* and *Thermos* are other notable examples.

5   I just have to tell this one! While living in Mexico, I happened to pick up a plastic bottle of soda which had the proud legend blazed across the label: 'Same drink! Different label!' Wow. Sign me up for some bulk-buying.

Monopolistic competition is like an endless game of catch-up; as there are no barriers to entry, successful products will immediately have a following of copy-cat products which will steal market share from the original manufacturing firm. The response is to differentiate the product as much as possible. Just follow the on-going battle between iPod and the myriad (= numerous) close-followers in the electronics industry – which is changing so quickly I'm not even going to bother with examples.

In the final analysis, the more consumers feel a product to be substantially different, the more successful the firm will be. I often claim that the standard catch-phrase of TV commercials 'New and improved!' would actually be a bit of an overkill if the firm actually meant it. It is quite sufficient for the firm simply to *convince* consumers that the product is different. One example rather close to us is the market for economics textbooks in the IB, where the first tailor-made textbook was by Alan Glanville, the second by me and the latest is the official IB version by Dorton & Blink. While these are indeed close competitors, there are distinct differences between them as we all seek to differentiate our books from the others. [6]

A Little Depth

My friend Toni took the following below from a sample out of the McGee watch collection – amidst snotty remarks from both him and his wife Pati that I should sell a few and buy a house (and get a ~~wife~~ life) instead. From left to right you see: Sinn (model Flieger 356 II) for USD1,500; Fortis (model B-42) for USD2,500; Breitling (model Chrono Avenger) for USD3,500; and a Panerai (Luminor GMT PAM88) for USD7,000. All are waterproof to between 100 and 300 meters. All are automatic (= self-winding). All have sapphire crystal glass. All are well within official chronometer limits (a maximum of +6 to -4 seconds deviation per day). Why are they here in the section

of monopolistic competition? Simple; they all contain the exact same mechanical movement ('ebauch' in watch-speak), namely the ETA-Valjoux 7750.

ETA is the largest manufacturer of Swiss watch movements in the world, and is owned by the Swatch Group. It supplies movements for the brands owned by Swatch (Omega, Tissot, Longines, Swatch to name but a few) and also sells to at *least* 100 other watch manufacturers – four of which are pictured above. Watch manufacturers making some of the most expensive mechanical watches in the world (IWC for example) to some of the cheapest (Aristo) all use the same movements from ETA. What differs is the price, obviously, but also the way the products are *differentiated*.

So what does the USD 3,500 Breitling do that the USD1,000 Sinn doesn't? Well, not much, I shamefacedly admit. Here's how the manufacturers of the four watches above sell their wares:

- **Sinn**: "…company started by legendary pilot Helmut Sinn…noted for advanced technology…advanced water-proofing using argon-filled chambers and copper-sulphate anti-humidity capsules…used by German special forces…"

- **Fortis**: "…aeronautical watches used by air forces… official watch of the European Space Station and Russian Cosmonauts…"

- **Breitling**: "…Instruments for professionals…officially certified by the Official Swiss Chronometer Testing Institute…" (Ads run with John Travolta standing in front of an aircraft, caption; 'Profession: pilot. Career: actor.')

- **Panerai**: "…originally made for the Italian navy commandos…unique patented crown lock system… highly limited editions…official suppliers to the 2006 North Pole Expedition by Mike Horn…"

So does the differentiation work? Well, I have one of each, you tell me. Am I the only sucker (= easy target) out there?

---

[6] Since Glanville has basically been in the business since the 1990s and helped write several IB syllabuses, his book might be considered the Yoda of IB econ literature. Dorton & Blink have the official version, e.g. Luke Skywalker. Mine? Probably Jar Jar Binks.

Summary & revision

1. The **assumptions for a firm operating on a monopolistically competitive market** are 1) Large number of suppliers with relatively small market share; 2) Products differ in some way, e.g. there is product differentiation; 3) The firm has a degree of market power due to its ability to depend on a degree of brand loyalty; 4) There are no or low barriers to entry.

2. The firm in monopolistic competition will be able to earn an **abnormal profit in the short run only**. Due to low barriers to entry, other firms will enter and steal market share. The market entry and exit process leaves the monopolistically competitive firm with **normal profits only in the long run**.

3. Monopolistically competitive firms will engage in **non-price competition** by way of product differentiation and branding.

32. Monopolistic Competition vs. Perfect Competition

Key concepts:

- Monopolistic competition and efficiency – comparison with a perfectly competitive market firm

Monopolistic competition and efficiency

Just as the PCM firm, the monopolistically competitive firm receives normal profits in the long run. However, as soon as one compares the two markets in terms of efficiency, the similarities end.

Figure 32.1, diagram a., shows that a firm operating in monopolistic competition with other firms, will set price above marginal cost (leading to suboptimal allocative efficiency) and produce at an average cost above AC_{min} (meaning that the firm is not productively efficient).

a. PCM firm

- **P = MC**; optimal allocative efficiency
- **Q = AC$_{min}$**; optimal productive efficiency

b. Monopolistic competition (LR)

- **P$_{LR}$ > MC**; suboptimal allocative efficiency
- **Q$_{LR}$ ≠ AC$_{min}$**; suboptimal productive efficiency
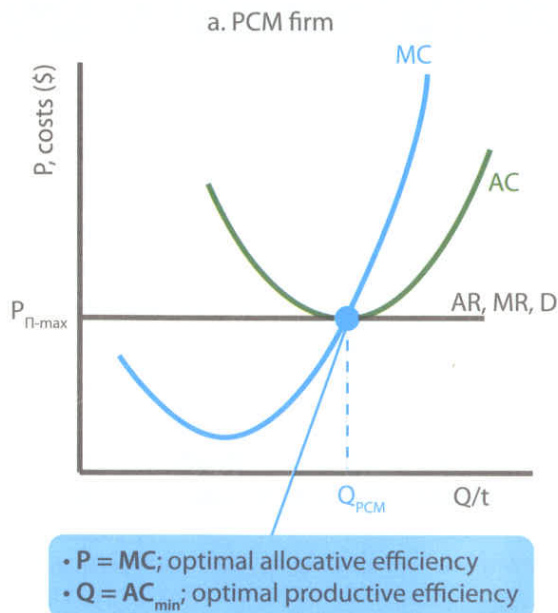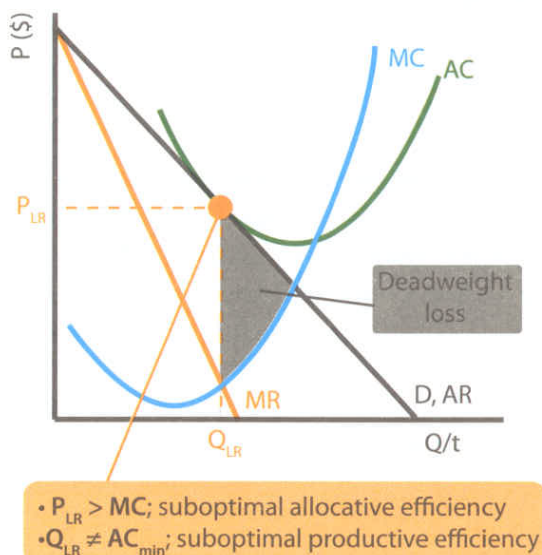
Figure 32.1 a&b Efficiency in monopolistic competition – a comparison to the PCM

The monopolistically competitive firm will produce at an output level **where neither allocative nor productive efficiency criteria are met**, i.e. the firm will not produce where P (AR) = MC, or at AC_{min}. [Note that suboptimal resource allocation can be seen in the **deadweight loss** above, the grey triangle in the right diagram.]

A degree of consumer sovereignty in the end

It would be far too simplistic – not to mention incorrect – to say that monopolistic competition is closer to a pure monopoly in terms of efficiency, or that market power enhances suppliers' power at the expense of consumers'. On the contrary, one may well argue that by continuously augmenting (= adding to) and differentiating products, the market creates a wide variety of goods and therefore enhances consumer choice far more than in a perfectly competitive market. Who's to say that the suboptimal allocative and productive efficiency is not worth this added choice? It is evident that the vast array of new products is quite often the result of non-price competition, which often takes place in monopolistically competitive firms.

Summary & revision

1. In the long run, the monopolistically competitive firm will operate at **normal profit** level.

2. Since the price will be set above marginal cost (P > MC), the monopolistically competitive firm will be **suboptimal in terms of allocative efficiency**.

3. The firm will not be producing at the lowest possible average cost ($Q \neq AC_{min}$). The firm is **not productively efficient**.

4. A **benefit of monopolistic competition** is the range of products, innovation and the addition to consumer welfare from a broader choice of goods.

POP QUIZ

Monopolistic Competition

1. What is 'imperfect' about monopolistic competition?

2. Explain how/why monopolistically competitive firms in LR must earn normal profits.

3. Using the concept product differentiation, explain how a monopolistically competitive firm might earn an abnormal profit.

4. What, if any, are the advantages of monopolistic competition over perfect competition?

5. Examine the level of efficiency in the long run in monopolistic competition.

33. Oligopoly

Key concepts:

- Assumptions of the oligopoly model
- Concentration ratio
- Interdependence
- Game theory; prisoners' dilemma

HIGER LEVEL

MICROECONOMICS

The final market version to be found lying in-between the extremes of perfect competition and monopoly is another form of non-price competition; oligopoly, meaning 'few sellers'. The defining line between monopolistic competition and oligopoly can be very thin, and many of the markets have in fact elements of both. Oligopoly markets are increasingly common in the real world – and perhaps the most important market structure to study. Unfortunately it has perhaps the weakest theoretical base as there is no single model which is able to conclusively describe it.

For a spoof on BOAC…and possibly collusion, see 'Comply with me!' at http://www.youtube.com/watch?v=e_mJdxtnpTQ asdf

Assumptions of the oligopoly model

An oligopoly has some key defining characteristics, namely that the market is dominated by a few large firms and that potential entrants face high barriers to entry – the latter being a key difference to the monopolistically competitive market. Furthermore, products can be both homogeneous and differentiated – for example both cement (homogeneous) and breakfast cereals (differentiated) are oligopolistic markets. With these outlining assumptions in place the oligopolistic market takes on a number of rather unique features which subsequently define the market.

Concentration ratio

Standard economic theory defines an oligopolistic market as one where the four or five largest firms control a major share of the market. This is the market **concentration ratio**, and is most frequently calculated as the total sales of an individual firm divided by total market sales. The rule of thumb is that when the four or five largest firms have over 50% of the market, an oligopoly situation is at hand. Note that while a highly

concentrated oligopoly of four firms might account for 90% of sales, the remaining 10% could be in the hands of hundreds of small firms. In some notable cases, two firms control 100% of the market, such as Boeing and Airbus on the market for large passenger jets. This is known as a duopoly.

Spot the duopoly

Oligopolies are found in all markets, from primary goods (oil, cement, aluminium), secondary goods (pharmaceuticals, automobiles, soft drinks), to tertiary goods (banking, telecommunications, air travel). The table in Figure 33.1 looks at a selection of four-firm concentration ratios (4FCR) in the US, in 2002.

Figure 33.1 Four-firm concentration ratios for various industries in USA (2002)

Industry	Four-firm concentration ratio
Automobiles	87.3%
Breakfast cereal	78.4%
Fibre optic cables	66.9%
Plastic bottles	42.3%
Daily newspapers	22.1%

(Source: 2002 US Economic Census of Manufacturing: www.census.gov)

One final statistic shows the extent of oligopolistic concentration on a global scale. The three largest soft drink manufacturers in 2010, Coke, Pepsi, and Dr Pepper Snapple, in 95 countries (which comprise 90% of the world's population) account for 85% of the market.[1] Two of the most concentrated global markets in the world are cigarettes (4FCR = 98%) and washing machines and dryers (4FCR = 90%).

Interdependence

A common denominator in oligopolistic markets is that there are very high barriers to entry, with well-established firms being able to heighten these barriers by using extensive advertising and branding to dissuade potential entrants. This is accentuated by the sheer size of many oligopoly firms, which are able to move along the LRAC curve and enjoy benefits of scale. This helps explain why many highly efficient firms in due course are left with few competitors, since firms which do not attain such scale benefits will be competed out of the market – or be forced to merge with another firm.

In the last decades of the 20th century, many large firms increased competitiveness by **merging**, which is the joining of two or more companies into a single firm. This furthered their ability to enjoy economic benefits of scale in advertising, bulk-buying etc. The merger process resulted in further consolidation (= bringing together, strengthening) of oligopolies, in that the 'few and large' controlled increasing portions of the market.

Having only a few *main* rivals to take into consideration, one would think that an individual oligopolist might be able to largely disregard them. Wrong. In fact, quite the contrary; any action taken by one major firm will have an immediate repercussion on the others. As a single firm will not have total control of the market but still be large enough to affect output and price, firms will be acutely aware of each other and will make it their business to have a great deal of knowledge about the others. This leads to a large degree of **interdependence** as each firm knows that any action on its part will cause some form of reciprocal (= 'in return') action or reaction; 'What goes around, comes around'. This makes it very difficult to predict price and output outcomes, as we shall see.

Game theory; prisoners' dilemma

Jack and Jill decide that carrying water up and down the hill is a waste of resources and decide to resort to the old-fashioned method of theft[2]. They illegally siphon water from a nearby reservoir and ultimately the 'leakage' is noticed. The police arrest them on charges of grand-theft but have only enough evidence to convict them of stealing a hose. The two are put into separate holding cells, all the while pleading their innocence to the crime of water-siphoning, but admitting to the crime of stealing hoses. The hose theft carries a two year jail sentence, and the interrogator knows that the prosecutor needs a confession to convict them of the greater crime of water-siphoning. The two accused are unable to communicate.

The crafty interrogator makes the following offer – separately – to Jack and Jill: 'We have evidence enough to convict you for the theft of the hose, and put you each in jail for two years. However, we know that you are also guilty of siphoning water. If I get confessions from both of you, we will give you each two years in a soft, minimum security prison. If one of you confesses while the other denies, the confessor will get one year but the other will get a 10 year penalty where you will spend your days with rapists, murderers, Ramstein fans and hardened criminals.'[3]

He then leaves the both of them to think over their options, which are:

1. Cooperate and stick to denial! This will result in two years in prison for each of them.

2. Confess and get a maximum of one year in prison.

As both Jack and Jill have two options each, there are four possible outcomes. Note that Jill's decision will affect Jack's outcome and vice versa! That is, the two are interdependent. Figure 33.2 shows the four possible outcomes in a payoff matrix.

Figure 33.2 Payoff matrix for Prisoner's Dilemma

---

1 http://www.beverage-digest.com
2 I am alluding to a childhood nursery rhyme; 'Jack and Jill went up the hill to fetch a pail of water…' and I forget the rest.
3 Don't send me angry letters – I'm a big Ramstein fan! *Du Hast* is playing as I write this.

Each one ponders away in his/her cell. The dilemma facing each of them is not knowing what the other will do! Let's look at Jill's options. If she sticks to her denial, she can hope to have a maximum of two years in prison – IF Jack also denies. Jill's temptation to confess comes with the realization that Jack too is thinking the same way. Thus, if Jill denies she runs the risk that Jack confesses, which lands Jill in prison for 10 years! The question both are asking themselves is 'I don't know what the other is going to do! Do I dare deny and put my hopes in the other also denying?!'

Collusion – i.e. both denying – would produce the optimum total outcome of two years each, but the problem is that Jack and Jill are not allowed to communicate. The best outcome for one of them is confessing – cheating on each other – while the other denies. If both decide to minimize the impact of not knowing what the other will do, then both will confess. The dilemma creates a movement towards the grey box in the payoff matrix above, an outcome known as a *Nash equilibrium*, where each player has adopted a strategy giving him/her the *best possible outcome given the expected action of the other*.

The oligopolist's dilemma

The line of thinking illustrated in the prisoner's dilemma can be applied to oligopoly strategy. Consider two oligopoly firms which have 100% of the market together, a duopoly. Posit that each firm initially is making a normal profit. Each firm wants to increase its profit and looks at the options. By raising the price in concert (collusion) both firms might make an abnormal profit as they would be acting as a monopoly. However, Firm A could stand to lose customers and profit if Firm B keeps the same price. If Firm A lowers the price it could make an abnormal profit – or the action could force Firm B to also lower the price and both lose! The final uncertainty of the other's action could force each firm into doing nothing – which would then be the best decision taking into account the expected action of the other – there would be a Nash equilibrium.

I have taken great artistic license and grossly oversimplified the use of game theory in the preceding example, and I am bending the syllabus boundaries as it is. There are many possible outcomes in a 'duopoly game', since the game would not be a 'one shot only' game such as portrayed in the prisoner's dilemma. Firm A will face the knowledge that firm B could retaliate A's price by lowering their price, leading to re-retaliation (?!) by A… and a price war results which is harmful to both firms. The outcome is most uncertain. Non-collusive oligopoly is often characterized by brief spells of competitive flurries yet broad price rigidity.

Summary & revision

1. The **assumptions for an oligopoly** are that the market is dominated by a few large firms on a market with high barriers to entry. Goods can be both homogenous and differentiated.

2. Market concentration is measured by a **four- or five-firm concentration ratio**. This measures the total market share of the four or five largest firms as a percentage of total sales in the industry.

3. The four or five major firms in an oligopoly are **interdependent** – e.g. the actions of one large firm in terms of pricing and/or product differentiation will cause a counteraction by the other main firms.

4. Game theory attempts to explain the possible strategic options available to firms operating under interdependent conditions. The most famous example is the Prisoners' Dilemma.

34. Collusive and Non-Collusive Oligopoly

Key concepts:

- Open collusion (cartels) and tacit (informal) collusion
- Non-collusive oligopoly
- The importance of non-price competition

'There are two paths you can go by, but in the long run…'

This is an excerpt from a classic Led Zeppelin song, and is relevant (for once, you probable say) in the context of oligopoly choice. Firms will know full well how 'tit-for-tat' competitiveness is potentially damaging to all firms, the extreme scenario being a full-blown price war where everyone – but the consumer! – loses. For example, consider a **duopoly** (two firms controlling the market) of two airlines operating on the same routes, which is rather common in many countries. Here are some of the strategic issues facing the two airlines:

- If one airline lowers ticket prices then the other might be forced to compete and lower its price…and both firms stand to lose revenue and profit.

- If one airline runs an advertising blitz to 'steal customers' from the other it might succeed – but the other airline might answer in kind to get its customers back. Both firms could spend money on costly advertising and ultimately wind up back where they started.

- One airline might enhance flight quality by offering pre-flight snacks, better on-board meals or TV screens at every seat. The other airline would have to counter…

- The two airlines could collude by setting the price of tickets or by dividing the market up so they never met in head-on competition on the same routes.

> **Definition: 'Collusive and non-collusive oligopolies'**
> A **collusive oligopoly** is when firms in an oligopoly market collude (cooperate) on prices, output or sales venues in order to maximise profits. If oligopoly firms instead compete independently one speaks of a **non-collusive oligopoly**.

The two paths the firms in the example can go by are basically to *collude* or *compete* (a non-collusive oligopoly). They could also simply do nothing! This is not as silly as it might sound. In the long run, either of the options will lead to a degree of *price rigidity*. Let us look at each option in turn.

Open collusion (cartels) and tacit (informal) collusion

When I was living in Sweden and had clean air, I ran marathons. Every year or so I had to get a new pair of running shoes and I would actually put it off since I knew how angry the process would make me. You see, the sport shoe retail outlet business in Sweden consists of a three-firm oligopoly which I am convinced are colluding. I base my suspicion on the fact that while the large brand names such as Nike and Adidas might be sold in all three outlets, I was never able to find the same model of shoes in all three places. This makes comparison shopping impossible – which is probably exactly the point![1] Retailers can avoid price competition by not competing with identical products – here,

---

1 One of my several ex-wives daintily avoided coming with me as I would inevitably vent my spleen (= unleash my anger) on the poor sales people. It was not a pretty sight.

simply by not carrying the same models.[2] **Collusion** takes on several forms:

1. **Open (overt) collusion:** If the main actors on the oligopoly instead have a *formal* agreement as to price and/or output, then collusion is open (sometimes called *overt*). This is basically a cartel and most countries have strict regulations against this type of competition-limiting behaviour; basically open collusion is illegal.[3] (See *Case study: Unhealthy cartel* below.)

2. **Tacit (covert) collusion:** There could be *tacit* (= unspoken) collusion where there is a price leader that the rest follow, or firms could follow benchmark prices such as recommended prices set down (legally!) by a producer organisation. Firms which have similar cost pictures could use the same mark-up (as in cost-plus pricing) and arrive at the same price level.

In any shape or form, collusion tends to move the market towards a monopoly outcome with very little price fluctuation as the price is more a result of agreement than market forces.

Ways of collusion

Firms have quite a few options when it comes to colluding, the most common forms being:

*   Setting the *price* is perhaps the most common method, yet there are many other ways to lower the competitive forces which harm oligopolist's profit. (Pharmaceutical companies)

*   *Dividing the market* into regions basically creates regional monopolies. (Glass companies in the EU)

*   Agreeing on *quotas* will limit supply and drive up the market price. (OPEC)

Collusive agreements seldom last for long and there is a simple explanation. Any type of collusion requires a manageable number of firms controlling the market and continuous coordination and openness between firms in order for all members to 'walk in step' and not break the agreement. There is also very little honour amongst thieves! Since the agreement

is illegal, there is no way for members to exert pressure on anyone who does not follow set prices or quotas. It is possible – even probable – that a member 'cheats' on the agreement in order to earn an even greater profit at the expense of members who stick to the agreement. All it takes is for one firm to sell above set quota or below set price for the collusive agreement to break down and the oligopoly to become competitive. Price will then tend towards a quasi-market solution with, again, little tendency for price change.

A collusive agreement taken to the extreme results in a **cartel**. A cartel acts to increase profits by setting/controlling output, or price, by dividing the market up between firms. The most famous cartels today are de Beers diamond cartel and the Organisation of Petroleum Exporting Countries – better known as OPEC. The aim is to avoid competition and act as one – a monopoly outcome is the end result. For this reason, cartels are illegal in virtually all industrialised countries. de Beers and OPEC are not subject to this restriction as they both to a large extent are based on an arrangement between governments and producers where governments have controlling interests. Also, the international aspect of these cartels transcends (= goes beyond) national law.

Cartels are notoriously difficult to uphold and maintain. The primary reason has been given earlier; the temptation for members to cheat, knowing that the arrangement is illegal in the first place and in the second place that one can earn additional profits at the expense of the others.[4] It is also hard to get all members to agree to a common set of quotas and/or prices, as these will be set with the common good in mind rather than that of an individual firm.

---

2   This is going to get censored but I'm still going to include it. I once compared oligopolies with men in cultures where virginity is highly prized; "…both seek to enhance their own competitiveness by avoiding good comparison shoppers…"

3   There is some seriously scary reading available at http://ec.europa.eu/competition/cartels/cases/cases.html

4   You need to be some kind of *stupid* to take notes of meetings, keep logs and put to paper/email all the activities and agreements involved in keeping up an illegal cartel. This is gleefully enough exactly what happened in 2007, when the European Commission found brewers Heineken, InBev, Grolsch and Bavaria guilty of running an illegal cartel for beer in The Netherlands during the period 1996 - 1999. Ample evidence of hand-written notes from meetings provided the basis for the verdict – and a rather hefty fine of €273 million (circa USD360 million at the time) for the brewers. I wonder if the brewers were perhaps…inebriated.

Story Time! Unhealthy cartel

In 2001, the European Commission fined eight companies €855 million for taking part in a cartel for vitamins and related products. The cartel had been operating since 1989 and had several distinct components; a market-sharing component where members were allotted certain products and a pure price-fixing cartel for vitamin products. The Swiss company Hoffman-La Roche was at the head of the cartel and as such received the highest fine: €462 million.

The Competition Commissioner for the EU, Mario Monti, called the cartel '... the most damaging series of cartels the Commission had ever investigated ...'. The commission found that 13 companies from Europe and abroad participated in the attempt to eliminate competition in the vitamin market, amongst them Hoffman-La Roche (Switzerland), BASF (Germany), Aventis (France) and Solvay Pharmaceuticals (Netherlands).

The EU commission found that the cartel was established high up in the hierarchies of the companies, pointing to a long-term strategic plan of dominating the global market for vitamins.

The cartel had regular meetings and a formal structure of management to exchange information on sales and production volumes. Output and revenue was carefully monitored in order for members to comply with quotas set by the cartel management. Amazingly enough, the same players had pleaded guilty to identical illegal collusion in the US in 1999! La Roche paid $500 million and BASF paid $225 million in fines.

As a striking footnote, the last year during which the cartel for vitamin C was in place brought total revenues of €250 million. Three years later – with the vitamin C cartel broken up – revenues were down to €120 million. This illustrates the lure of cartels quite clearly.

I also have to comment on the size of the fines. However difficult it is to estimate the gains of collusion, one can look at the total revenue to get some idea of the severity of the fine.

Hoffman-La Roche had sales of €10.8 billion in 2001, which means that the €462 million fine was just under 4.3% of sales. The EU commission can set fines of up to 10% of sales.

(Source: http://www.health.fgov.be/WHI3 , www. roche.com & www.eu-oplysningen.dk/euidag) Non-collusive oligopoly

Non-collusive oligopoly

'Business is a good game – lots of competition and a minimum of rules. You keep score with money.'[5]

One way to illustrate the aforementioned outcome of price rigidity in non-collusive oligopoly situations is the **kinked demand curve**[6].

Consider a firm operating in an oligopolistic market, where a non-collusive price is (somehow) arrived at; PEQ in Figure 34.1 below. Just as in the prisoner's dilemma, the firm is in a highly interdependent situation, where a change in price will certainly lead to a reaction from other firms in the oligopoly. The model suggests the individual firm will arrive at the following; 'If I raise my price nobody will follow since they will expect to steal my customers. If I lower my price everybody will follow in order to hinder me from stealing their customers!' Plotting this out in Figure 34.1:

- A *higher* price will cause a proportionally larger decrease in revenue for the firm and thus the demand curve will be *elastic* above the starting equilibrium price of P_{EQ}. The firm would lose market share to other firms by raising the price.

- In *lowering* the price, a firm would quickly be followed by competitors who themselves would fear losing sales to the price-lowering firm. Therefore any decrease in price below P_{EQ} would mean a more *inelastic* demand curve and a decrease in revenue.

The demand curve will be 'kinked' at the point of equilibrium price and output, and as the MR curve is dependent on the shape of the demand curve, the MR curve is discontinued.[7]

---

5 Statement by *Nolan Bushnell*, founder of Atari. Taken from *Perloff*, page 419

6 Not 'kinky' but kinked, meaning 'bent'.

7 The issue of what the MR curve should look like in the kinked demand curve is subject to heated discussion. When the first edition came out, my friend Alan "Gandalf" Hobson immediately took issue with the fact that I'd left a portion of the lower part of the MR curve above the Q-axis – while claiming that the relevant portion of the demand curve was inelastic. Alan, I've thought it through. You are right. Another discussion took place during the Oxford Revision Courses in 2010 when I had a furious discussion with my dear friend George...in a pub in front of some pretty terrified colleagues. The lesson is; NEVER discuss economics over Vodka! My other dear colleague Konstantin has since made sure that he doles out the Stolichnaya a bit more sparingly. Yes, it was your fault Konstantin.

Elastic demand

Inelastic demand (note that since only the inelastic portion of the D-curve is used, MR will be negative)

P, costs ($)

P_{EQ}

Q_{EQ}

D, AR

MR

D, AR

MR

Q/t
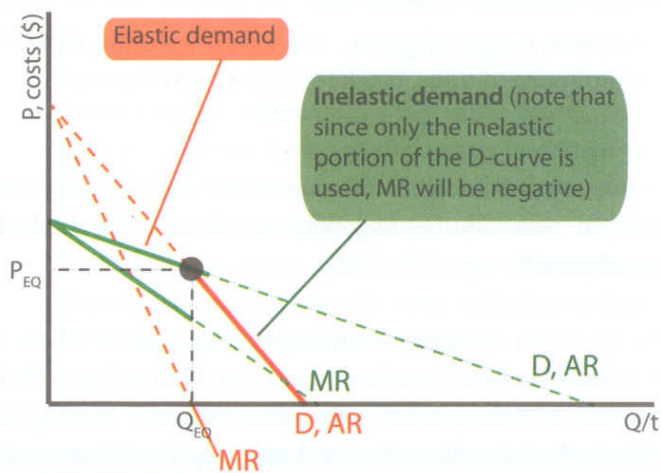
The kink in the demand curve means that the MR curve will be discontinuous – broken.
The dotted extensions of the curves are only to show how each MR curve will cut 'halfway' on the Q-axis. (Don't include these in your own diagrams!)

MC_0

MC_1

MC_2

Q_{EQ}

P, costs ($)

D, AR

Q_{EQ}

MR

Q/t
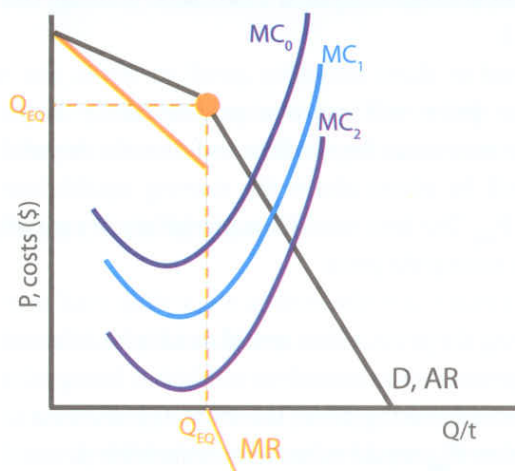
The broken MR curve enables a **range** of possible MC = MR points. The diagram shows that marginal costs can change without the oligopolist having to change output in order to stay at the profitmax level of output.

Figure 34.1 Kinked Demand Curve

We are assuming the firm to be a short run profit maximiser, and the discontinuous MR curve gives a range of possible profitmax points for the MC curve to pass through. This doesn't mean that output has a range, but that the firm has a number of possible MC curves anywhere between MC_0 and MC_2, enabling absorption of any additional costs without the output level being affected. In other words, an increase in marginal costs would not necessarily have any effect on output or price since the firm could remain at Q_{EQ} and simply accept a lower profit margin.

The model of the kinked demand curve helps to explain two key characteristics of oligopolistic markets:

1. Why oligopolies have a tendency towards *price rigidity*. The 'kink' in the demand curve is the same situation as in the prisoner's dilemma; a Nash equilibrium, where each firm's decision to leave the price untouched is actually the best outcome for the firm – as long as all the other firms act the same!

2. Why oligopolies so often do not compete in price but rather resort to various forms of *non-price competition*. (See next heading.) Any form of price competition amongst the few large firms controlling the market can well lead to **price wars** – e.g. where one competitor lowers prices and others retaliate by lowering even further... creating a downward price spiral which ultimately only leaves one winner; the consumer.

It bears mentioning that there are conspicuous *weaknesses* in the model:

- Firstly, it doesn't explain how the *equilibrium price* is arrived at in the first place.

- Secondly, dynamic market effects – supply and demand changes – resulting from *non-price competition* are not explained.

- Thirdly, the model does not allow for any *other interpretation* of the demand curve and how firms would act if their view of the demand curve turns out to be incorrect.

The importance of non-price competition

Price competition is evidently something that is not a major factor in the competitive processes of oligopoly markets, but rather a situation to be avoided. Firms will instead seek to enhance their products by differentiating and profiling their product in other ways than via price, thereby avoiding head-on competition. Here are a few examples of how an oligopolistic firm can compete other than via price.

Quality and innovation

In an argument similar to that in the monopoly model, large oligopoly firms could well have large abnormal profits available for R&D, quality enhancement and innovative new products. Using, once again, my passion for wristwatches as a starting point, Rolex has a long history of innovation; patenting of the perpetual motion mechanism (a self-winding watch mechanism); the famous 'Oyster case' was made in 1926

thereby creating the worlds' first truly water-proof watch; a date was added in 1945, resulting in the Date-just watch (virtually unchanged today it is probably the most recognised watch in the world); the Submariner watch became the first diving watch in the late 1950s and so forth. The latest version of the Rolex Sea-Dweller ('Deep Sea') is waterproof to…wait for it; *3,900 metres*[8]! Yes, it's on my wrist as I write this.

'Inventing a better mousetrap' is indeed an incentive for the firm as this will serve to *differentiate* the product from others on the market. For example, auto manufacturers will put inordinate effort into quality control and will continuously strive to increase fuel efficiency, performance and safety. Retail chains can present higher (perceived) quality by offering longer opening hours, delivery service, on-line ordering, and advantageous credit purchases. Oligopolistic producers often seek to create a number of different versions of a product for different market segments – e.g. Toyota, which produces economy cars under the Toyota label and luxury cars under the Lexus brand – in order to make demand as inelastic as possible for each single product.

Enhancing perceived value

One of the world's most famous marketers, Philip Kotler, uses a model of 'Core, tangible and augmented product' to explain how firms can develop a product's profile[9]. A firm in non-price competition will seek to add to the perceived value in consuming the good by focusing on many possible intangible aspects – the 'feel' of the good.

The 'core' of the product is about what the consumer is looking for in the product. It is the basic perception of utility that the firm tries to create; a car is not 'transportation' but a 'driving experience' for a BMW owner and 'sophisticated comfort' for a Cadillac owner. The core product must be backed up by any number of 'tangible' (= concrete or physical) qualities such as packaging, styling and design, brand image and special features; the BMW has an aerodynamic form, alloy wheels and a spoiler package as optional extras, while the Cadillac may have a cigar humidor, wine cooler and Sensuround TV built in.

Finally, the producer attempts to 'augment' (= increase, amplify) the buyer's sense of benefit by stressing the additional post-purchase value of the good; BMW has free service for the

first three years and Cadillac has a 24 hour breakdown service including towing, repairs, and replacement vehicle.

Branding and advertising

In a manner of speaking, brands originated during the middle ages when guilds (= associations for craftsmen) forced craftsmen to put a trade mark on their products in order to guarantee a certain level of quality for the consumer – and probably to keep out unwanted competitors! Oligopolistic firms of today put great effort into dividing up the market and focusing the marketing mix on specific target segments. Volvo sells cars using the 'safety' argument; Toyota is 'reliable and economical'; while BMW uses 'performance' – these marketing efforts serve to differentiate the products and at the same time limit the degree of head-to-head competition.

Marketing and advertising is very costly. Famous-brand companies will pay millions for a 30 second TV commercial during, say, the Oscar Awards. These costs can actually serve to create barriers for firms which seek to enter the market – knowing how difficult it will be to compete with high-profile brands which have had years to enhance brand loyalty amongst customers. The marketing costs serve to create a 'one-and-only', 'high-quality' and/or 'luxury' image for the good. It must, however, be continuously upheld in order for the producer to exact a premium price for its products.

An increasingly common method of advertising is product placement in films by making sure that the product is clearly visible in a scene or two. In actual fact, sometimes the product takes on a roll central to the film! Just look at the most successful series of movies ever, the James Bond movies. How many products can you think of which have been firmly associated with Bond? Aston Martin and BMW in cars; Stolichnaya and Absolut Vodka in drinks; and Rolex, Breitling, Seiko and Omega in watches since no James Bond movie would be complete without the wristwatch gadgetry supplied by 'Q'.

"Well Moneypenny, I must be off. So many ~~women~~ evil-doers and so little time!"

8 www.rolex.com and *Wristwatches* pages 350 - 361; *Brunner, Gisbert L. and Pfeiffer-Belli, Christian*, Germany, 1999, ISBN 3-8290-0660-8

9 *Kotler, Philip; Marketing management*, Prentice-Hall International Editions, 1988, sixth edition; ISBN 0-13-556267-8

'What's in a name!?' goes an old saying. The answer in connection with branding is 'A lot!' Brands have become increasingly valuable assets for firms. For example, Interbrand – a renowned branding consultancy firm – values the above-mentioned Rolex brand at close to $5 billion. The top three global names in term of brand value in 2008 were Coca Cola ($66.6 billion), Microsoft ($59 billion) and IBM ($59 billion).[10]

And yet the small survive

In view of the on-going process where (global) oligopolies arise, how do so many small firms manage to remain? Part of the answer is to be found in human nature. People tend to feel comfortable with the close and familiar, so a local fast food place will have its place in spite of any number of international chains. It also makes it easier to get additional service, say in repairs or follow-up service – anyone who's bought a bicycle knows the value of local service! Another reason for small scale success is that many services do not easily benefit from scale, for example haircuts, shoe repairs and garden services. Finally, local tastes, customs and preferences make large scale 'streamlining' more difficult for large firms and also give small/local firms an edge in quickly adapting to changes in local tastes. In short, it is a challenge for firms to be both global and local.

Summary & revision

1. A **collusive oligopoly** is a market situation where oligopoly firms cooperate by way of setting prices, output or dividing the market between them.

2. A **non-collusive oligopoly** is a market situation where oligopoly firms compete and each firm is acutely aware of the other firms' actions – there is a high degree of interdependency.

3. **Open collusion** is when firms form a cartel, e.g. a formal organisation formed by firms in order to set price, output or in other ways manage the market.

4. **Tacit collusion** is when firms follow a 'price leader' or some form of benchmark prices rather than a formalised pricing structure.

5. Open collusion can be done by agreeing on the **price**, setting **output quotas**, and **dividing the market** up geographically.

6. Oligopolies are often characterised by **price rigidity** and **non-price competition**. The **kinked demand curve** helps explain this to a certain extent.

7. Methods of non-price competition include focusing on **R&D** and **innovation, enhancing perceived value** and **branding** via heavy advertising.

---

10. http://www.interbrand.com/best_global_brands.aspx

35. Price Discrimination

Key concepts:

- Definition and types of price discrimination
- Reasons for price discrimination
- Pre-requisites for price discrimination
- Advantages and disadvantages of price discrimination

When I was a young man and frequently out being naughty, I was often infuriated by a policy of discrimination found in discos and clubs where men paid an entrance fee and women did not. It was not until many years later that I understood the underlying economic logic in the price-setting: more women mean more men, and since men apparently have a more inelastic demand for clubbing (and for women?) they can be charged more than women. I said I understand it – I still don't like it. Perhaps it is a good eye-opener as my feelings might be something similar to those that women have endured for ages. Just a thought![1]

Definition and types of price discrimination

Price discrimination means that different people will be willing to pay different prices for the same good. The next time you take a really cheap flight or bus/train trip, try to get into a conversation with a well-dressed individual sitting in the waiting lounge. If it turns out that this is a businessperson on a business trip, tell him/her that you are an economics student and that you are writing about pricing policies of airlines/ buses/ railroads. Ask what he/she paid for the ticket – I wager that it will be a considerably higher price than yours! The businessperson has been subjected to **price discrimination**, since the airline/ bus company has sold the same basic good at different prices to different consumer groups; a business person who is willing/able to pay a relatively high price and a student with higher price sensitivity for tickets[2].

---

1 Favourite in-class fun with my crowd; ask the students if they have been out clubbing. There are always a couple who have. Pick one of the lads and ask him if he paid. Ask one of the girls if she paid. Then explain: "Josh, you paid a bundle of cash to get into the club and Jasmine got in for free. Right? Here's why Josh; your demand for her is *much* greater than her demand for you! Thus, your PED for getting into the club is much lower than Jasmine's."

2 Be prepared for some anger. Many people do not realise the

> **Definition: 'Price discrimination'**
> When firms actively adjust the price of a good according to the willingness/ability (e.g. according to the price elasticity of demand) of different consumers to pay for it, the firm is price discriminating.

Firms operating in perfectly competitive markets are price-takers and therefore lack pricing power. A firm with pricing power, e.g. a monopoly, can set any price – within the confines of demand – as it is operating behind entry barriers of some sort. Thus far we have assumed single-pricing strategies in our models, yet nothing prohibits a firm from charging *different* prices for a good. Economic theory identifies three main ways to price discriminate:

1. **First degree price discrimination;** the firm charges whatever the market demand will bear. Auctions and 'haggling' at open markets are a form of price discrimination, as each good goes to the highest bidder.

2. **Second degree price discrimination;** the firm sells at quantity discounts or 'tiered markdowns'. For example, many publishing houses will have different prices on school textbooks depending on the quantity ordered. The first 30 copies cost $25 per book, the second thirty $22 per book, and any order above 60 receives a 15% discount.

---

extent to which they have been discriminated against. A fellow passenger once caused quite a scene when he found out that I paid less than 10% of what he had paid.

3. **Third degree price discrimination;** the firm separates the market into distinct groups or segments and sells the good at different prices. The disco club and flight ticket examples fall into this category – as does most price discrimination. We will deal solely with 3rd degree price discrimination here.

Reasons for price discrimination

There are three main reasons (plus one *Outside the box*) why firms would have an incentive to price discriminate:

1. The most obvious reason for price discrimination is that selling a good at different prices will serve to *increase profits*. This is intuitively obvious; a firm that has market power can increase profits by selling the good at the highest possible price according to consumers' willingness to pay. The firm might also use price discrimination to cut prices in order to force competitors out of the market (predatory pricing) or in order to penetrate a new market.

2. Firms might also think in long run terms, and price discriminate in order to *capture a future market*, for example by letting students purchase magazine subscriptions at a lower rate, the firm can look forward to higher future subscriptions – at the normal subscription rate. An export firm selling at a lower price abroad is very common and such practice is often vilified (= spoken ill of). This form of discrimination helps exporting firms attain economies of scale and strengthen competitiveness. In addition to the above pecuniary (= financial) motives, one should not disregard the goodwill firms can earn by letting certain groups benefit from lower prices; senior citizen discounts on buses; student access to educational material on the Internet; free parking for handicapped, etc.

3. All the above examples arise from the private sphere, yet a good many examples of price discrimination arise in the arena of *public and merit goods*. Public utilities can charge lower rates to firms than households in order to provide competitive infrastructure for industries. Price discrimination can also help public monopolies to cover costs in high cost areas, for example by discriminating in railroad tickets the additional costs of providing train services to rural areas can be covered. A frequently used form of public price discrimination is based on concepts of fairness, where it is considered beneficial to society that all groups have equal access to goods. Such

discrimination is often based on *income*; high income earners could pay more for municipal child care and schooling than lower income households.

4. Finally, as an *Outside the box* issue, it is possible for a price-discriminating firm to supply a good where a single-price market simply would not exist, due to high average costs. Discrimination can also allow a public monopoly such as a utility service to produce at an even *higher output level than the break-even* level of output (i.e. beyond the point where AC = AR).

Applied economics; innovative price discrimination

One of the most inventive forms of price discrimination I have heard of in later years came from my friend Glenn. He runs a ju-jitsu and his dojo faced the same profit issues as other gyms; fluctuating demand and therefore highly uncertain revenues. Glenn's solution was as brilliant as it was simple; charge differently for old and new students.

The basic idea is to charge an entry level price which never changes! A new student in 2011 would pay SEK 2,000 (about $250 at 2003 exchange rates) for a year-long membership – while students who joined the club in 2005 continue to pay the membership fee they started with. The SEK 2,000 fee will never change as long as the student continues to be a member. However, as soon as a member leaves the club and doesn't pay that year's membership fee, he/she will start all over again if/when a new membership is bought. Thus, a student who entered in 2005 would still pay SEK 1,800 in 2003, but by leaving the club during 2009, he/she would pay the (higher) entry-level fee upon re-joining in 2011.

This is quite brilliant. It provides an incentive for students to remain in the club which diminishes drop-out behaviour. The club will keep more of its students and have a much better grasp of future costs and revenue.

Methods of price discrimination

There are any number of ways for a firm to price discriminate. The most common are:

1. **Time;** airline tickets are subject to 'X weeks in advance' and movie theatres have different rates for different times of the day. Electricity, squash court rental and fishing licenses will all be fairly easy to divide into 'peak' and 'off-peak' groups of demand.

2. **Age;** apart from previous examples, there are under-12 discounts for cinema tickets, amusement parks and air travel[3]; while senior citizen discounts are often available for such services as haircuts.

3. **Income;** lawyers often price discriminate for their services according to income as do prostitutes and private tutors[4]; health care and college tuition are also often cases where different income groups pay different prices.

4. **Gender;** the aforementioned discrimination of nightclubs against men is one example.[5] Other examples – against *women* – would be haircuts and personal hygiene products.[6]

---

3 Notice that airline tickets are used as an example in many methods of price discrimination. I strongly suspect that the confusion as to what an airline ticket actually costs is most intentional. Airlines simply don't want price transparency as this would lead to price competition. They want us confused, and by using several layers of discriminatory pricing, airlines increase our search costs of finding correct price information. Yet the only thing I want to hear is 'Smoking or non-smoking, Sir?'

4 I love putting these two professions in the same bracket. Guaranteed to yank a few chains in the halls of learning.

5 I keep hearing from my students at the OSC revision courses how their teachers have said that this is in fact *not* an example of price discrimination but rather a marketing method – more women attract more men etc. Yes, agreed. However, there is nothing in the definition of price discrimination which stipulates that there are given reasons *why* firms might price discriminate. The basic fact is that when a firm charges different prices for the same good it is price discrimination – regardless of the underlying motives for the discrimination. Yes, I looked it up.

6 Yes, I looked this up too. It seems that women's PED for products dealing with personal cleanliness and appearance is lower than men's! It works for me: I shocked my students and colleagues in an interview in the school paper by commenting on how shampoo and dish-washing liquid were perfectly substitutable for me. Since dish-washing liquid is cheaper I use that to wash my ~~hair~~ head.

Pre-conditions for successful price discrimination

There are three pre-conditions for successful price discrimination.

1. The first is that the firm must be able to identify *different distinct groups* of consumers. As an extension of this, the groups must have distinct differences in their price elasticities of demand. Otherwise the firm would be unable to charge different prices in the first place.

2. The second requirement is for the firm to be in a *position of market power*, i.e. be able to set prices, which excludes the possibility of price discrimination in a perfectly competitive market as firms operating under such circumstances will not have pricing power. I will use a monopoly to illustrate the effects of discriminatory pricing.

3. The final and rather obvious condition is that the firm must be able to *limit the effects of arbitrage* – i.e. re-selling. If I could get one of my students to buy my airline and cinema tickets at student-discount prices I would. This must be rendered impossible by the producer in order to be able to charge some groups higher prices. Thus, airline tickets have the travellers' names on them which must match their passports, while the student-discount cinema ticket can only be used upon showing a valid student ID. Basically, firms must be able to keep the different groups separate in order for the discriminatory scheme to function. This doesn't necessarily have to involve complicated administration in separating the groups, since there are often built-in segmenting walls, for example when an international firm sells services and price discriminates between countries.

Figure 35.1 below shows how a price-discriminating monopoly arrives at the profitmax price and output for the two market segments I and II, each with its own PED. The sum of MR ($MR_I + MR_{II}$) for each segment equals total market MR, and profitmax output at MC = MR is 40,000 units. The differing elasticities of demand in market segments will set two different prices; $7 in market I and $5 in market II. (Note that the diagrams only illustrate how output and price is set – not the amount of profit. This is shown in Figure 35.2 - price discriminating monopoly – further on.)
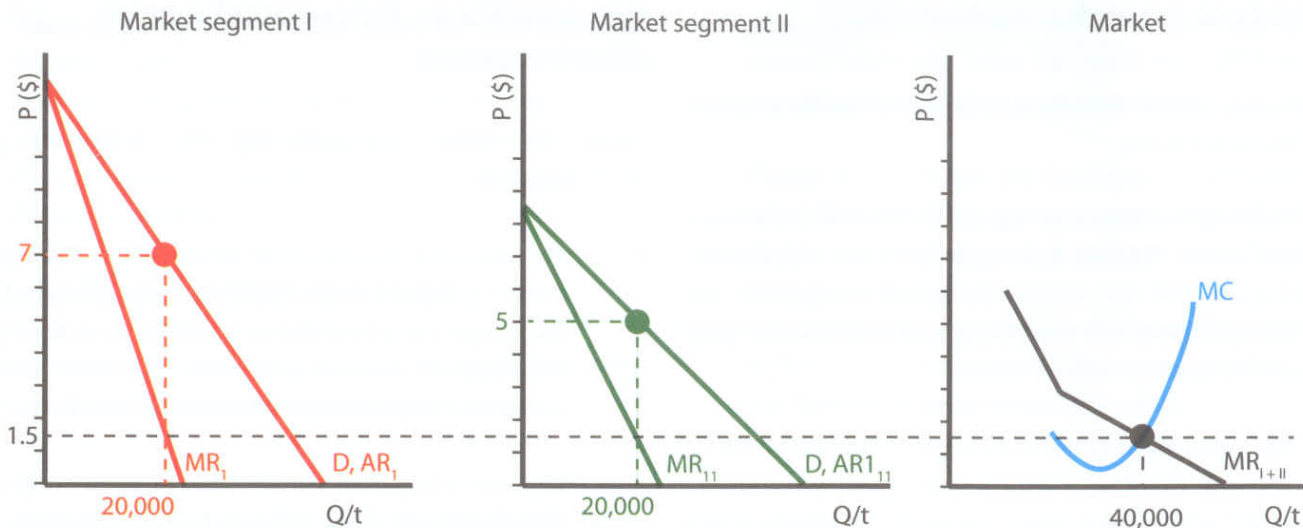
MICROECONOMICS   HIGER LEVEL

Market segment I

Market segment II

Market

Figure 35.1 Price discrimination – duo-price monopoly

Market effects of price discrimination

Now we look at how producers and consumers can stand to gain and/or lose by the process of price discrimination. Assume a monopoly firm setting a profitmax price and output, shown in Figure 35.2 below. (Please compare this to Figure 35.2, as it builds on those figures.) The *single-price monopoly* is making an abnormal profit of $120,000 ($3×40,000 units) at the profitmax point of output. Now, assume that the firm is able to identify two distinct groups having different demand and sets the highest possible price for half of total output. Diagram B shows that these 20,000 units could be sold at a price of $7. The *dual-price monopoly* charges two prices and sells 20,000 units at $5 and 20,000 units at $7. Now, here's the tricky bit; as output has not changed there is no change in costs! This means that average costs are the same, $2, having been set by the original output of 40,000 units. Therefore the element of abnormal profit per unit goes from $3 to $5, bringing the profit for the higher-priced batch to $100,000. Adding the profit from the lower-priced batch ($3×20,000 = $60,000) brings the total abnormal profit up to $160,000. The net addition to abnormal profit by price discriminating is $40,000, shown by the square in diagram B.
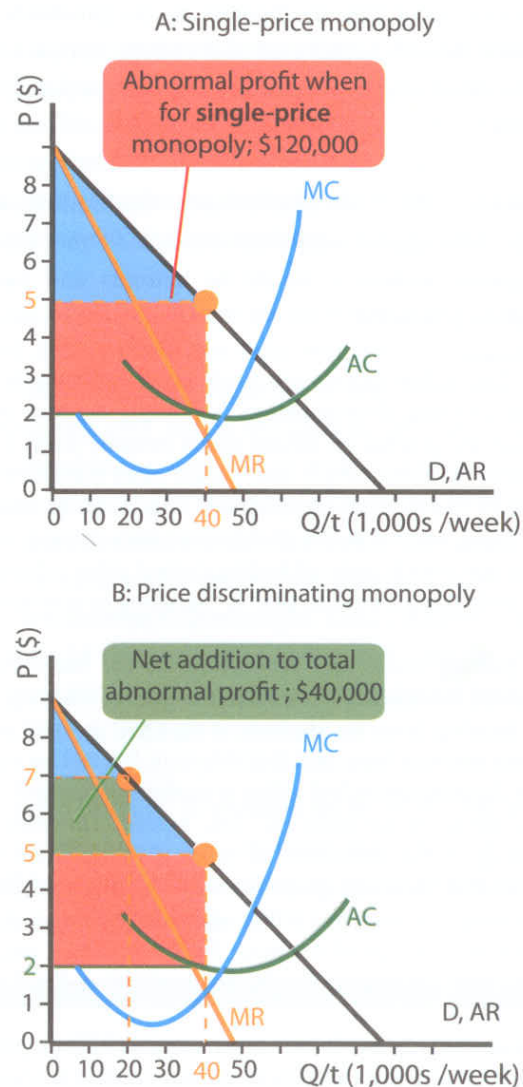
A: Single-price monopoly

B: Price discriminating monopoly

Figure 35.2 A price discriminating monopoly

A Little Depth

Further price discrimination

Let's continue one step further. Say that the firm manages to divide the market into four segments, where each group can be priced according to its demand, shown in Figure 35.3. The highest-demand group will pay $8 for 10,000 units; the second-highest pays $7 for 10,000 units and so on. The addition to total abnormal profit by further price discrimination (going from two prices to four prices) is $20,000, shown by the two striped squares in diagram C. By dividing the market into four price groups, the firm makes a total abnormal profit of $180,000.

C: Further price discrimination

Net addition to total abnormal profit ;
$10,000 x 2 = $20,000
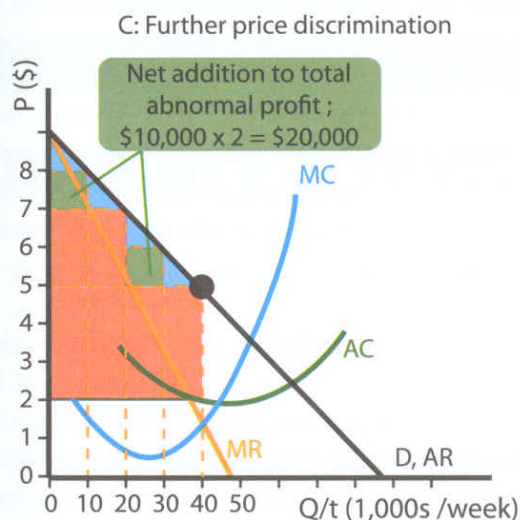
By ever-finer division of the total market into segments, the price discriminating firm increases total abnormal profits.

Figure 35.3 Normal profit and loss in monopoly

Outside the syllabus: Note that the finer the demand segment, the more consumer surplus is captured by the firm. Remaining consumer surplus is shown by the increasingly 'saw-toothed' blue triangles in the diagrams.

Advantages and disadvantages of price discrimination

Producers

The advantages to the producer are obvious; higher total profits. As an extension of this, a producer would be able block market entry by setting output at the break-even point (AC = AR) and make up for this by selling 'portions' of total output at discriminating prices – the producer would still be able to make an abnormal profit in spite of producing at break-even point. One can say that the abnormal profits gained by price discrimination serve to pay for an amount of goods sold at break-even cost. This will dissuade potential entrants as it will be most difficult to match the incumbent's price.

Consumers

Consumers stand to both lose and gain from price discrimination. Some groups will benefit from lower prices while others will 'subsidise' the winners. Clearly consumers would lose if competition fell due to predatory pricing by the price discriminator, but again, abnormal profits allow the firm increased funding to plough back into the business in the form of R&D, resulting in new products and improved quality. Yet the main objection by consumers is probably the transfer of income to the producers, since the overall picture is that consumer expenditure increases but not necessarily the quantity consumed. The transfer, or 'capture' of consumer surplus by the producer clearly shows the redistribution effects. On an international scale, there is a bit more conclusive evidence of consumer loss, since multinational companies often price discriminate on a country-wide basis owing to the fact of pure geographical barriers to arbitrage. This helps explain why – ceteris paribus – international firms often set the price higher in high-income countries than in low-income countries, for example medicines, paperback books and services.

Society

It is also possible that both consumers and producers gain from price discrimination. One possible advantage to society due to price discrimination is the possibility of producing a good where normally no firm would be willing or able to supply it. No firm would produce in a market where AC was higher than the profitmax price. Figure 35.4 illustrates the price of round trip travel by car on the bridge between Denmark and Sweden; at the profitmax (= lossmin!) output of 4,000 units the demand curve renders a maximum price of SEK 500 – but the average cost is SEK 550. It is impossible to make a profit – the loss is SEK 200,000.
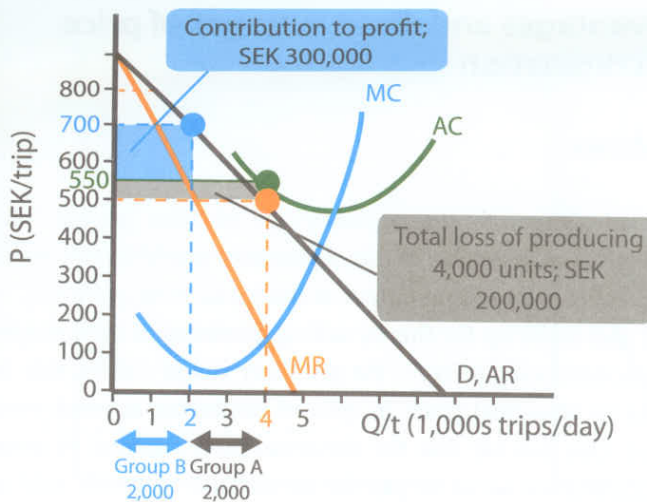
Figure 35.4 A Loss-making monopoly; Öresund bridge

A price discriminating firm could counteract the loss by setting output at profitmax and then discriminating against consumer groups A and B by, for example, setting a higher price for peak-time travel. 4,000 units are produced at an AC of SEK 550. 2,000 units are sold to **group A**; the total loss is SEK 50 × 2,000 = SEK 100,000.

The remaining 2,000 units are sold to **Group B** – e.g. people willing to pay a higher peak-time price – at a price of 700. (Remember that total output sets the 'floor' for AC!) The profit per unit of the units sold to Group B is SEK 150, giving a total of **SEK 300,000** as a contribution towards making up for the loss of SEK 200,000. In the final outcome, **total net abnormal profit** is the total profit contribution minus the total loss; SEK 300,000 – SEK 200,000 = SEK 100,000.

A Little Depth

Possible re-distribution gains

Price discrimination is often practiced by publicly-owned monopolies, such as mail service, telecommunications and railroad travel. Price discrimination stands to benefit consumers by enabling a monopolistic firm to break even – or even earn a profit - from a non-economically viable activity which would normally not be undertaken.

The aim of public monopolies is often to provide a societally optimal level of output rather than a profitmax level. The monopoly could seek simply to break even in terms of profit and set output where average costs equal average revenue – 70,000 units in the diagram.
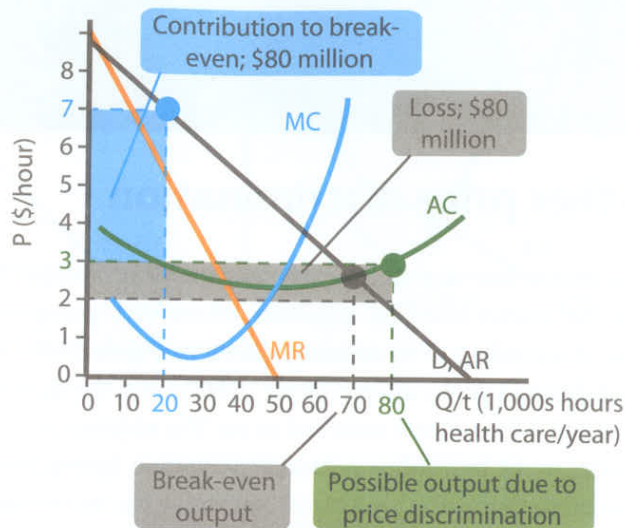
Figure 35.4 B Price discriminating monopoly

However, posit that installing telephone lines, railroad tracks and postal routes to outlying rural districts comes at a very high marginal cost. Should these members of society pay for this or should society as a whole pay? If an economy chooses the latter – as many in fact do – then the social benefits of supplying all areas are considered to outweigh the costs. Yet price discrimination can actually enable the public monopoly to cover the higher costs.

Assume that the public health service monopoly wishes to provide all areas in the country with equal amounts of health care hours. Initial total quantity provided is 70 million hours per year, but an additional 10 million hours of health care is needed to give what is considered uniform health coverage throughout the country. Producing 80 million hours of health care will cause average costs to exceed average revenue, resulting in a loss of $80 million (grey area in the diagram) at an hourly rate of $2. The monopoly can make up for this loss by charging higher rates for certain groups, such as rates based on income or higher rates for non-essential (cosmetic) surgery.

The diagram illustrates a possible (highly idealised – as in 'mermaidomics') outcome, where discrimination- pricing at $7 per hour adds a contribution of $80 million, resulting in a break-even outcome for the monopoly. One can say that the 25% of the population paying a higher price have subsidised the lower cost of health care for the remaining 75%. The discriminating monopolist has benefited society.

POP QUIX 35.1

1. A firm charges a higher price for its product in a certain geographical area due to higher storage costs. Why is this not price discrimination?

2. What difficulties would arise for firms trying to price discriminate on sales made over the Internet?

3. How does the possibility of arbitrage (re-selling) affect a firm's ability to price discriminate?

4. Outside the syllabus! Assume that a firm could *perfectly* price discriminate, i.e. set a separate price along every possible point on the demand curve. Where would profitmax be? What would happen to consumer surplus? And, really tricky question; where would the MR curve be?

5. Explain how a society might actually benefit from price discrimination. Account for both 'winners' and 'losers'.

Summary & revision

1. Price discrimination occurs when firms with market power set different prices for the same good according to different price elasticities of demand amongst consumers.

2. Economics identifies **three types of price discrimination:**

   a. **First degree** price discrimination: firms charge whatever the market will bear and attempts to get each individual consumer to pay whatever he/she is able and willing.

   b. **Second degree:** this is when firms set per-unit prices lower when the buyer purchases larger quantities – a bulk-buyer discount.

   c. **Third degree:** when sellers identify different groups of buyers and sell goods according to the PED for each group.

3. Firms price discriminate for several reasons. The obvious reason is that since average costs do not change the can increase profit by selling some of the units at higher prices. Other reasons include **market capture** by a firm wishing to **expand and/or attain economies of scale;** earn goodwill by setting prices lower for disadvantaged groups; public monopolies might price discriminate in order to increase output to a more societally beneficial level; a loss-making monopoly can use price discrimination to recoup some/all of the loss.

4. There are many **ways for a price discriminating firm to segment** the market into groups. The more commonly used segmentation variables are age, sex, time, income and geographical location.

5. The **necessary preconditions** for successful price discrimination are: 1) There must be distinctly **identifiable groups with different PED** values; 2) The seller must have a **degree of market power;** 3) The seller must be able to **limit arbitrage** – e.g. groups getting the good at a low price selling to groups prepared to pay a higher prices.

6. When sellers divide the market into ever smaller segments, **profit for the seller rises and consumer surplus decreases.**

7. Possible **advantages for firms** of price discrimination are increased **profits**, possibility of increasing **market share** and then attaining **economies of scale**, and **societal goodwill**.

8. Possible **advantages for consumers** include **lower prices** for some groups, **improved products** if firms use increased profits for R&D, and – in the case of a loss-making monopoly – **getting the good produced at all**.

9. Possible **disadvantage for firms** is that separating the market increases administration costs.

10. The potential **disadvantages to consumers** would be that some consumers end up paying a **higher price** (and they will not necessarily be the poorest groups), the **loss of consumer surplus**, and increased abnormal profits might help finance **predatory pricing**.

11. A **publicly owned monopoly** can increase societal welfare via price discrimination by setting the price at a loss making level (MC pricing) and then recouping some/all of the losses by charging certain groups more for the good/service.

2.1

36. Introduction to Macroeconomics – Economic Activity and the Circular Flow Model

Key concepts:

- Introduction to macro; main macro goals
- The circular flow of income model
- The identity of $E \equiv O \equiv Y$

In this section we take a few steps back from the individual markets we have looked at so far in order to get an overview of all the different markets in an economy; goods, services, capital, exports, imports and labour. These are **aggregates** of individual markets and the relationships between them are, rather unsurprisingly, highly complex resulting in a great deal of disagreement amongst economic theoreticians as to how things actually work. There are, however, several points which economists are able to agree on with a minimum of bloodshed, namely that there are a number of overall **macroeconomic goals**. These goals can be encompassed in a 'Four plus three' list as follows:

Introduction to macro; main macro goals

Growth – defined as the increase in gross national product (GDP) or gross national product/gross national income (GNP/GNI) measured in money terms and adjusted for inflation

Price stability – defined as a stable incremental increase in the price level (inflation) and measured by the consumer price index (CPI[1]) and the GDP deflator.[2]

Low levels of unemployment – defined as the number of people, out of the total available workforce, who do not have jobs and given as a percentage value.

External equilibrium – defined as stable exchange rates and balance between imports and exports (this is the focus of Section 4).

In addition to the above four 'mainstream' goals, economists increasingly bring up issues to be found within the following captions:

Environmental concerns – often defined as economic growth in compliance with 'non-depleting' and 'non-degenerative' resource use and measured using various environmental indicators of pollution over time

---

1 Another measure of inflation commonly used is the retail price index (RPI).
2 There are a number of other indices designed to calculate 'underlying' or 'core' inflation. More on this later.

Distribution of income – defined as how well income is spread between the richest and poorest sections of the population.

Productivity of factors – defined as the output per unit of input (factors of production) and often measured by labour productivity indices and capital use per money value of output.

All the above are subject to a variety of government policies aimed at achieving the goals or targets. The problem is that it is impossible to achieve all of the goals at once. In fact, virtually every macroeconomic goal will conflict with at least one other goal. A simple illustration is when an economy experiences increased output (goal 1 above) this often leads to inflation (making goal 2 more difficult) and possibly an environmental burden (conflicting with goal 5). As we shall see in Chapters 57 and 59, these conflicting goals will create numerous trade-offs for governments which in turn adds to an already heated theoretical and political debate.

The circular flow of income model

Let us expand the scale of our economy a bit and assume that it is comprised of households and firms – no financial institutions, government or other countries exist. The flows involved are shown in Figure 36.1.

Firms create the output (**O**) which is consumed by households – this is expenditure (**E**). (Note that since only a household sector exists, all expenditure (**E**) is in fact consumption expenditure (**C**).) Households are the actual owners of production factors and rent these out to firms for which an income is received (**Y**). (It is imperative that you view firms as separate entities which are owned by households! Ownership is either directly or via shares in the company. In other words, while the firm is responsible for producing goods and services, the buildings, machines and labour are supplied by households to firms for a price.) National income is thus a **flow concept** since we are measuring how much money is passing through the system during a given amount of time (normally one year) to handle all the economic transactions.

In order to drive home the point that expenditure, income and output are indeed the same flows measured in different ways, I have compared the model to a closed-circuit heating system. (Yes, I know I'm being childish.) Picture a furnace in the basement of a house which heats water (see illustration) and forces a flow of hot water through a piping system and a heating radiator in a room above. The system is a closed circuit – no water can leak and no additional water can be injected. Therefore, no mater where one puts a 'probe' to measure the

flow of water per unit of time the amount would be the same. Just think of a river which has narrow and wide sections; the same amount of water per minute flows by regardless of width – assuming that there are no tributaries (= inflowing streams) or seepage into the ground.

> **Definition: 'National income'**
>
> National income is the real money value of the sum total of all final goods and services produced in an economy during a given time period – usually one year. This gives the identity of expenditure ≡ output ≡ income, in money terms. As income is measured as an amount between periods, it is a flow concept.
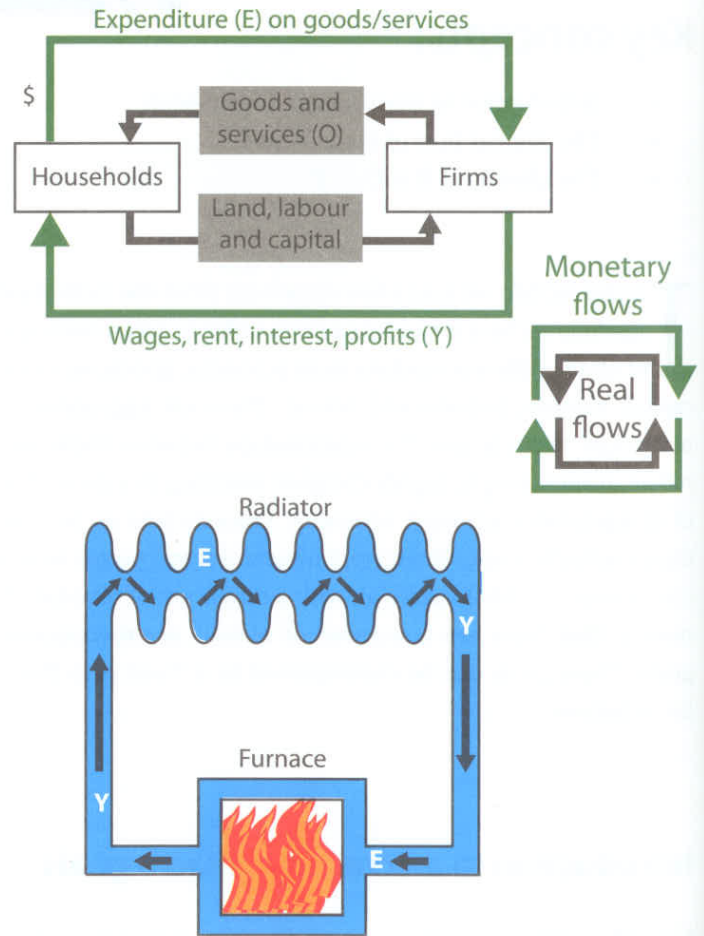

Figure 36.1 Circular flow in a simple economy

The flows in the simple circular flow above match each other since we are making a number of assumptions in this simple model; there are no taxes, savings or imports so households cannot do anything but spend their income. Also, there are no exports, government spending or investment, which means that firms cannot sell their goods/services anywhere but to domestic households. This provides an uninterrupted flow since there are no financial leakages or injections into the system. The model also shows that there are two parallel flows in operation. One flow (inner black circular flow-arrows in Figure 36.1) shows a **real flow**; land/labour/capital from households to firms corresponding to the output going from firms to households. The **monetary flow** (outer green circular flow-arrows) shows how households are rewarded with wages, rent, interest and profits (**Y**) from firms while firms are on the receiving end of households' expenditure (**E**).

Adding financial institutions, government and a foreign sector

We now relax the above assumptions and include other parts of the economy making the model more realistic. Let's follow the monetary flow of income through the system and see what happens. Starting at **Y** in Figure 36.2, a household receives an income of, say, $100. The household will not spend all of this on output from domestic firms:

- A portion of the income flow will go to the government coffers as **tax** (T);

- Some will go to financial institutions (e.g. banks) as **savings** (S);

- And some of the remaining income will be spent on goods produced in other economies, i.e. expenditure on **imports** (M).

Assuming that households pay $25 in tax, save $10 and use $15 on imports, then there are $50 left for domestic consumption (C). In other words, out of an income of $100 there has been a **leakage** (L) of $50 out of the system.
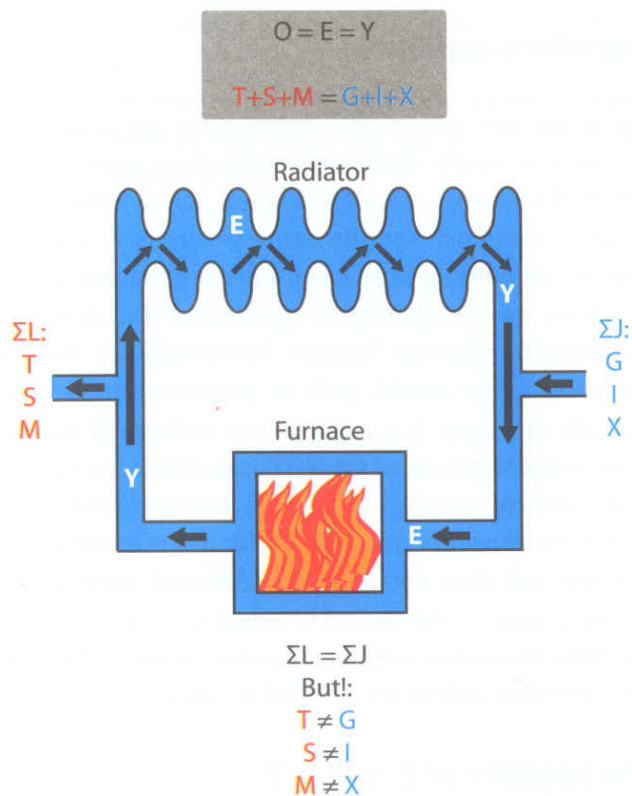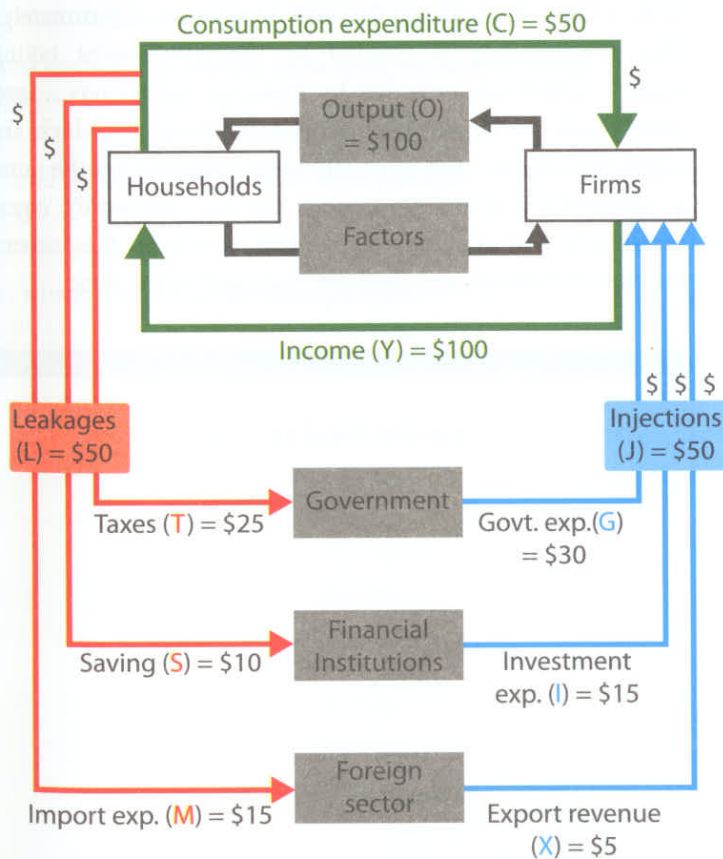
Figure 36.2 Circular flow in a more complex economy

What happens to this $50? Simple, it flows in from other sectors in the model – these inflows are called **injections** (J). (Referring again to the radiator model, there **has** to be an inflow of 50 or the water pressure will not be enough to keep the flow up.)

- The taxes (T) will provide governments with funds needed for hospitals, defence etc., the public sector. This is **government expenditure** (G) which flows back into the system, i.e. to the firms providing these goods.

- Households' savings will provide financial institutions with loanable funds which firms use for **investment expenditure** (I).

- And finally, there will be expenditure from the foreign sector which is an inflow of **export expenditure** (X) from abroad.

To continue with the example of the 'missing' $50, posit that government spending is $30, investment is $15 and exports $5. This gives a total inflow of $50, i.e. **total injections are equal to total leakages**, since it is a definitional impossibility for more money to be pumped into the system than ultimately flows out. In the final calculation, the money value of injections must equal leakages. The money value of output must equal total expenditure – which in turn is the same as income.

Transfer payments

You might have noticed that certain flows are missing in the circular flow model. Most governments have provisions to aid household incomes via some form of social welfare system, or social safety net: social security payments to low income households; housing allowances; student grants for university education etc. In addition to government transfers, certain economic transactions between households are missing in the circular flow model, such as payments for used goods and gifts of money. It is important to understand that all the above **transfer payments** are NOT included in the calculation of government spending in national income accounting. This money has *already* been accounted for in the national income accounts and does not represent additional expenditure, so adding it anew to the flow of G would be to double count it. The basic rule is that only the expenditure flows which have a **corresponding output** are included in national income.

The identity of E ≡ O ≡ Y

I usually introduce this keystone macroeconomic illustration by pulling a few coins out of my pocket (and as I never seem

to have any money I mostly have to borrow them from one of my people) and 'buying something' off the table of a student; for example, a time management calendar (my people are scarily well-organised) for 100 Swedish Crowns (SEK). This simple transaction illustrates the entire macro model in one fell swoop; I have spent SEK 100 (= expenditure, E) on a good valued at SEK 100 (= output, O) and the student has received the SEK 100 as a payment for his/her output (= income, Y). Assuming a 'two person economy' as above, we see that total **expenditure** equals **income** (in money terms) which equals **output**. As a result of how we define expenditure, income and output in national income accounting, the three flows resulting from the simple economic transaction above are per definition **identities**, i.e. $E \equiv O \equiv Y$. We have now measured economic activity by assigning a monetary value to it, the SEK 100 paid for the calendar. On a wider scale, national income is simply the sum of all such transactions taking place during a given period of time.

It is worthwhile to warn of a common error concerning injections and leakages. While it is quite apparent and perhaps even logically obvious that each leakage has a 'mirror image' or 'counterpart', this most assuredly does not mean that the values will be identical. The 25 money units of household taxes paid in the example above resulted in government spending of 30, i.e. G > T, an all too frequent occurrence unfortunately.[3] This overspending is enabled by the government taking loans – which are provided by financial institutions… and these funds are supplied by depositors' incomes…which are supplied by firms…and so on. Therefore taxes are not the same as government spending; savings do not necessarily equal investment; and imports are not equal to exports. It is instead the **sum** of these in- and outflows that must be the same.

The circular flow model renders two sets of identities:

$$O \equiv E \equiv Y \text{ and } \Sigma J \equiv \Sigma L$$

---

3 This is the government budget deficit, and causes government to borrow – creating government debt. More later.

A Little Depth

Regarding the flows on the left hand side in Figure 36.2 (Circular flow in a more complex economy) it is clear that the income received by households goes to tax, saving, imports (consumption of non-domestic goods) and domestic consumption. Thus:

$$Y = T + S + M + C$$

Now look at the right side of the circular flow model. The very same firms which are paying for the use of households' factors of production are on the receiving end themselves; consumption expenditure, government expenditure, investment expenditure, and export expenditure flow to the firms. We get:

$$E = G + I + X + C$$

And since $Y = E$, putting both sides together gives us: $T + S + M + C = G + I + X + C$

...and subtracting C on both sides gives us: $T + S + M = G + I + X$

... which is nothing else than: $\Sigma L = \Sigma J$.

Summary & revision

1. The **main macro goals** are *growth* (increase in real GDP); *price stability* (measured by the consumer price index, CPI); *low unemployment* (defined as the percentage of the total labour force not holding a job); and *external equilibrium* (stable exchange rate and balance between exports and imports).

2. Other objectives include *environmental concerns* (e.g. sustainable development); *distribution of income* and *factor productivity*.

3. **National income** (here GDP) is the money value of all goods and services produced in an economy during a period of time – usually one year.

4. The **circular flow model** consists of households, firms, government, financial institutions and the foreign sector.

5. **Injections** in the circular flow model are government spending (G), investment (I) and export revenue (X).

6. **Leakages** are taxes (T), savings (S) and import expenditure (M).

7. The sum of leakages is **equal** to the sum of injections.

8. The three ways by which we **measure national income** are expenditure (E), output (O) and income (Y). These are an identity; $E \equiv O \equiv Y$.

MACROECONOMICS

37. Measuring Economic Activity – GDP and GNP/GNI

Key concepts:

- Measuring economic activity – GDP and GNP/GNI
- Output, income and expenditure methods of GDP accounting
- From GDP to GNP
- Nominal and real GDP
- Nominal and real GNP/GNI
- Per capita income
- Use of national income figures
- Green GDP

Measuring economic activity – GDP and GNP

An electrician once took upon himself to explain to me – in between his bouts of drinking my beer and coffee while pretending to install a washing machine – how the flow of electricity through the wires resulted in different measurements such as 'ohms', 'watts' and 'amperes'. I never understood it, but it seemed clear enough that final output of electricity could be calculated by using a number of different parameters – all of which had very specific meaning. Calculating output/income/expenditure is much the same since each different method results in identical values. The rest of the chapter will show different ways of calculating GDP and subsequent adjustments to this base figure. The plan of attack here is in three steps:

- Compute the money value of aggregate output during a year, i.e. gross *domestic* product; **GDP**

- Adjust for the use of foreign factors of production to arrive at gross *national* product/income; **GNP/GNI**

- Take into consideration changes in the price level to show real output, *real* national income; **GDP**$_{real}$ and **GNP**$_{real}$

Output, income and expenditure methods of GDP accounting

1. Factor income method of accounting

Adding up total payments for use of factors of production paid to households is commonly divided into four flows of payments; income from employment (wages) and income from self-employment (wages and profits); income to firms (profits); rents; and interest. These four correspond to the four factors of production; labour, entrepreneurs, land and capital.

The easiest way to exemplify income accounting is to use some actual figures. The 2001 national income accounts for Ireland will be used for all three methods, starting off with the figures for factor incomes below, Figure 37.1. In all three accounting methods I shall comment briefly on the posts and figures compiled in the accounts.[1]

---

1   The official figures for national income vary enormously in Official Accounts from country to country and can be immensely difficult to plough through. I owe a large debt of gratitude to Margaret Power at the Central Statistics Office of Ireland (*An Phríomh-Oifig Staidrimh* in Irish) for assistance in putting the following figures together for me.

Figure: 37.1 Gross domestic product in Ireland 2001 – (factor) income method of accounting

Factor income	Amount (millions of €)
Employment income (wages)	47,090
Self-employment income (wages + profit)	10,903
Rental income	6,181
Private/public gross profits (profit)	26,290
Interest	11,619

Total domestic income

less stock appreciation	218
statistical discrepancy	569
GDP at factor cost	102,869

- **Employment income** is readily understandable; it is the flow of payments to providers of labour for their services, called either wages or salaries.

- **Self-employment income** is the income generated by own businesses. It is accounted for and contains a profit element since an owner-operated business will generate a value-added element which goes to paying the owner's wages.

- **Rental income** is generated when owners of land, housing and property receive payments for other economic agents' use.

- Private and public **profits** are commonly separated in the accounts, but I have lumped them together here since the concept is the same; surpluses created by firms which are then paid back to the owners. This can be done directly (in small companies) or indirectly via dividends (= payouts) to shareholders.

- **Interest** in national income accounts is the bank interest earned by households. Households' deposits are used by firms and the payment for the loans is the interest paid to households. (It is in fact net interest since households' interest payments are deducted.)

- In assembling all the millions of figures comprising the national accounts, one tries to be as true to real output as possible. This can be tricky when dealing with the money values of goods since – as we shall see – price increases will distort (inflate) real output figures. We deduct **stock appreciation** in order to use a truer value since unsold goods and half-finished products lying in warehouses for lengthy periods would be given a higher value when added to the GDP figures than when originally produced. For example, a firm produces 100 Widgets in June and stocks them at a list price of AUD5 each, at a total value of AUD500. During the time these goods sit on the firm's shelf waiting to be sold, the price level increases by 5% and when the firm's annual report is filed the firm can list the unsold stock as an asset worth AUD 525 since inflation has added an additional AUD 25. However, the real contribution is only the original AUD 500 and the stock appreciation must be deducted.

- The final figure is then adjusted for a **statistical discrepancy**, which is basically assessed by comparing the values in all three methods and adding/subtracting an error component arrived at by comparing to an average based on all three accounting methods.

The compiled figures give us **GDP at factor cost**, which means that the final figure is to the furthest possible extent based on real use of factors of production which will compute with real expenditure figures in the next method of GDP accounting.

2. Expenditure method of accounting

The basic accounting premise here is that all expenditure flows from economic agents (households, firms, government, and foreigners) constitute total expenditure and thus income. Basically, this method of accounting looks at total spending during a period of time and divides the spenders up into groups in order to follow the flows and see who is spending on what. Total expenditure in an economy becomes the sum of consumption expenditure (C), investment expenditure (I), government expenditure (G), export expenditure (X) minus the import expenditure (-M). While you skim through Irish expenditure in Figure 37.2, try to figure out why we subtract imports from GDP.

MACROECONOMICS

Figure: 37.2 Gross domestic product in Ireland
2001 – expenditure method of accounting

Expenditure type	Amount (millions of €)
Household consumption (C)	55,202
Total private expenditure on capital (plus physical change in stocks) (I)	27,461
Government expenditure (G)	15,413
Exports (X)	112,938
Imports (-M)	-95,702
less taxes on expenditure	-14,572
plus subsidies	2,697
statistical discrepancy	-569
GDP at factor cost	102,869

- **Consumption** measures the amount of personal money spent on goods and services during the year. Consumption is often divided into **durables** (cars, refrigerators etc.), **non-durables** (beer and Donegal tweed) and **services** (such as car repairs, hotel stays and banking).

- **Investment** is firms' expenditure on capital goods such as machines, equipment and factories, often referred to as **fixed capital formation**.[2]

- Governments build roads, hire more teachers and buy fighter aircraft from domestic firms – this is simply your tax money buying goods and services; **government spending**. Since it is far easier to account for the market price of a jet fighter than 100,000 school hours, many services are estimated at the cost of provision rather than market prices.

- If an Irish company sells €1 million worth of knitwear but domestic expenditure for this good is only €900,000, then there would be a discrepancy between the value of expenditure and output, i.e. E ≠ O. This explains why **export expenditure** – foreigners' spending on Irish goods – is added, since we are estimating the total expenditure on goods produced in the country.

- In adherence with this, any expenditure by Irish on non-domestic goods must be deducted. This **import expenditure** does not represent any domestic output and represents a flow of money out of the system. Often one sees the term 'net exports' used, which is the product of export revenue minus import expenditure.

- We now arrive at **GDP at market prices**, which must be adjusted for two systematic inconsistencies; taxes and subsidies. Since we are measuring expenditure, most goods will include a proportion of indirect (expenditure) *taxes*, such as value-added taxes (VAT) and excise duties. Since these taxes do not have any corresponding output they must be subtracted. *Subsidies* skew the figures in the opposite manner, since the value of subsidies lowers the final market price below the actual factor cost. Therefore subsidies must be added on.

After, once again, adjusting for statistical errors we arrive at **GDP at factor cost**, which is the same value as in the income method used earlier. These two methods show the 'flip sides of the coin', as income in an economy must equal expenditure, which in one of the most common formulaic expressions is; $Y = C + I + G + X - M$. We now finish with the method used to calculate the *physical* output in money terms.

3. Output method of accounting

In calculating GDP using money values of total output one must be aware of the possibility of *double counting*, i.e. take measures not to count the same good twice. The issue of double-counting is important enough to warrant a brief example. Let's follow a product, peat for gardening, through the process of manufacturing to final purchasing at an Irish gardening store. Assume at the initial end of the chain a landowner in County Cavan who has a few thousand acres of prime peat bog and at the other end of the chain a consumer who walks into the flower shop to buy a bag of garden earth.[3] To simplify the example, we

---

2 The total figure on investment also includes changes in stocks, *circulating capital*, since any unsold goods produced in the time period are part of inventory and still represent output even though they have not been sold. Unsold and unfinished goods are accounted as expenditure by the firm. Say a firm produces €100,000 worth of Widgets but sells only €90,000 worth. If we counted only the expenditure the firm has received then actual periodic output would be underestimated by the additional Widgets valued at €10,000 now in the firm's warehouse. This addition to stock must therefore be included in the accounts. Similarly, had inventory fallen by €10,000 then €10,000 must be subtracted, since total expenditure exceeds the actual amount produced in the time period. Circulating capital also includes the elements of completed work in long term projects, '**works in progress**', such as airports and roads.

3 I've received a few emails from students who had no idea what 'peat' is. It is basically decomposing vegetable matter ('wannabe-coal'?) formed over millions of years and often found in marshy wetlands. It can be used for building material, fuel and garden soil.

will assume that there are only four stages in the transaction chain:

Stage I: The landowner, Paddy, sells the unprocessed raw bog – stripping rights – to Paddy, at the Soggy Bottom Peat Company for **€10,000**.
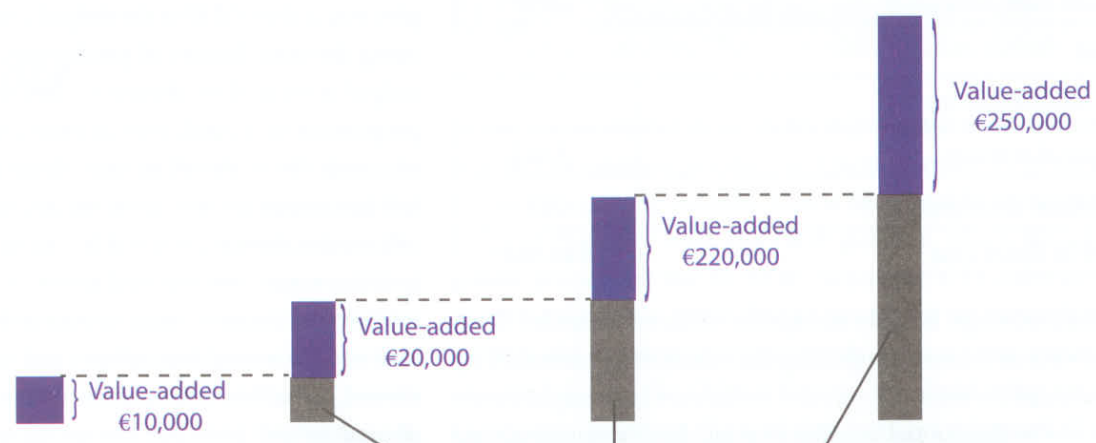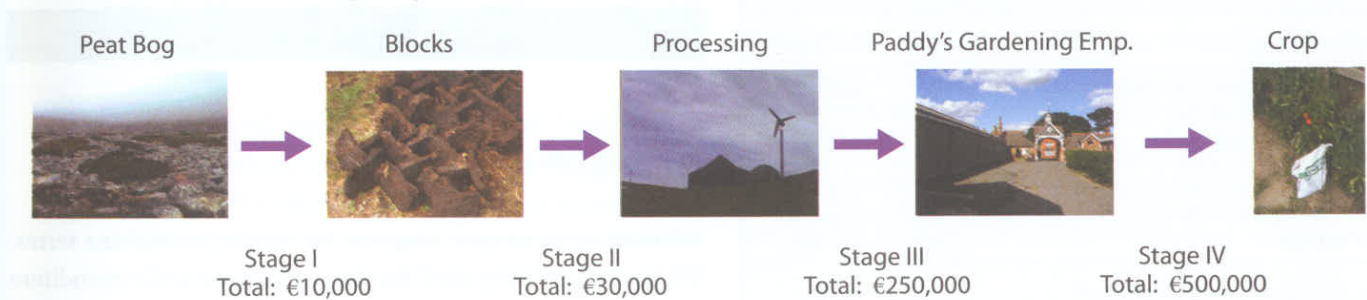
Stage II: Paddy at the Soggy Bottom Peat Company then cuts out cross sections of peat and dries it and sells it for **€30,000** to…

Stage III: … Paddy & Paddy at The Leprechaun Garden Boyos, a garden wholesaler. The peat is carefully bagged and labelled with 'Gnome-approved' stickers and sold for **€250,000** to…

Stage IV: … Paddy's Gardening Emporium, which, after sticking on labels reading 'Made from Gnome-friendly Irish peat!' are sold during the year for €10 apiece, totalling **€500,000** in retail sales. It is quite possible that some of the customers are named Siobhan (pronounced 'Shevaun'[4]).

---

4   Hey, don't look at me. While my forefathers came from Ireland, I doubt that we invented the spelling.

Trick question:

What is the total value of **final output**, i.e. GDP, in this 'economy'?! If your initial impulse is to start adding the value of output at each stage – i.e. €10,000 plus €30,000…etc – then stop immediately and read on instead. Adding the output at each stage of production is exactly what we are trying to avoid, since we would be double-counting values at every stage. Figure 3.1.5 shows how the initial output of peat bog sold for €10,000 is purchased by the peat company and sold on for €30,000. The value of output at Stage I is €10,000 and at Stage II €30,000. However, the original €10,000 has been included in the output value of €30,000 at Stage II when it is sold on to the garden wholesaler – this €10,000 must be deducted in order to see the value of output which is linked solely to the Soggy Bottom Peat Company, i.e. the value-added of €20,000. And so it continues on up the chain. Each consecutive stage buys the output of the previous link in the chain and adds value before selling it on. Total final value of output is the €500,000 paid at retail level by consumers – which is identical to the sum of all value-added through the chain; €10,000 + 20,000 + 220,000 + 250,000 = €500,000.

Peat Bog	Blocks	Processing	Paddy's Gardening Emp.	Crop
Stage I Total: €10,000	Stage II Total: €30,000	Stage III Total: €250,000	Stage IV Total: €500,000	

Value-added €250,000

Value-added €220,000

Value-added €20,000

Value-added €10,000

The expenditure on factor inputs used in production has already been accounted for and must therefore be removed when calculating the final value of output. Otherwise the same intermediate output would be counted several times over.

Figure 37.3   Value-added in output method

Calculating GDP using the output method is commonly done by collecting the figures on value-added for firms. In actual fact, it is virtually impossible to measure output by attempting to count 'final output' in every industry. Imagine an economy comprised of three firms; an iron mine, a tool firm and a building firm. The iron goes to the tool firm. The tools go to building firm....and the building firm uses the tool to build a shed for the iron mine! Oh, and the mine buys tools from the tool firm. So, how does one calculate final output of each firm without double counting?! The answer is, one summarises total output value (equivalent to total sales revenue) for each firm and deducts the costs of factor use. This shows how much value each firm has added and Figure 37.4 shows how GDP is arrived at using this method.

Figure: 37.4 Gross domestic product in Ireland 2001 – output method of accounting

Output type	Amount (millions of €)
Agriculture, forestry, fishing	4,003
Electricity, gas, water	1,329
Construction	8,085
Manufacturing	32,715
Transport, communications	6,014
Wholesale & retail trade	8,728
Banking, finance, insurance	8,478
Other services	22,789
Public administration, defence	4,032
Health care, education	9,991

Total

less adjustment for	
financial services	-3,863
statistical discrepancy	569
GDP at factor cost	102,869

The methodology is fairly straightforward; value-added from all the various sectors comprising the economy are summed up and adjusted for financial services (which are interest payments that must be discounted in order to avoid double counting) and a statistical discrepancy. We get the same value of GDP as in the previous two cases.

Finally, the importance of avoiding double counting brings me to one of the most common misuses of economic principles, namely the insistence of a good many laymen and even well-

regarded economists in comparing the *revenue* of multinational companies (MNCs) with the *national income* of selected countries. This muddled and rather misguided comparison is ostensibly an attempt to show the 'power' of MNCs in comparison with – and often put in terms of 'at the expense of' – less developed countries.

POP QUIZ

GDP Accounting Methods

1. Explain why the value of final output must be equal to the sum of value-added.

2. Why do we deduct transfer payments when calculating GDP?

3. Calculate GDP from the following figures: Consumer expenditure = €15 bn, general government final consumption = €8 bn, gross domestic investment = €6 bn, value of total exports = €3 bn, value of total imports = €4,5 bn, capital consumption (depreciation) = €2,5 bn.

From GDP to GNP

As usual, one can go a long way by simply 'tasting' the terms. We have so far computed the flow of income and expenditure created within a country, GDP, where the 'domestic' part should give you a clue. GDP is the output created within the economy using *domestic* factors of production. However, not all of this output is created by domestic companies since there will be a proportion of foreign-owned enterprises operating within the economy. In addition to this, domestic companies will hold foreign assets in the form of subsidiaries and joint-owned companies abroad. Both will lead to in- and out-flows of income (profits, wages, interest and rents) in each country. A country will receive income from property held abroad (subsidiaries and wholly owned businesses) and will pay property income abroad. By taking into account this **net property income from abroad** to and from the foreign sector, we get **gross national income, GNI - or gross national product, GNP.**

Definition: 'Gross national product (GNP)'

GDP is an account of the money value of goods and services produced within an economy – regardless of domestic or foreign ownership of the firms. **GNP** takes into account foreign ownership in the economy and domestic ownership of firms abroad by adding on net property income from abroad.

GDP + property income from abroad – property income paid abroad = GNP…or…

GDP + net property income from abroad = GNP

To summarise:

- Gross **domestic** product means 'produced within a country's boundaries'. These are goods resulting from home-based assets – regardless of whether they are owned domestically or by foreigners. GDP is thus delineated by country boundaries; 'Where (produced) – not who (owns)!'

- Gross **national** product on the other hand, means 'produced using a given country's factors' – regardless of where. GNP deals with ownership originating in a certain country; 'Who (owns) – not where (produced)!' In the next section, we will look a little closer at the question of which measurement constitutes the 'best' indicator of economic performance.[5]

Nominal and real GDP

As GDP is comprised of millions of different goods it would be virtually impossible to measure it in actual quantities of the goods produced which is why it is put in *money* terms. The problem with using money as a measurement is, of course, that the value of money continuously changes – there is *inflation*. One unit of a currency does not buy the same amount of goods if prices have increased. And since we measure GDP using money terms, the value of output has been inflated by the increase in prices because output is calculated by taking the quantity times the price for all units of output.

---

5   By now you should know that the answer will start with "It depends. Assume the following…"

Understanding the difference between real and nominal is very important in economics. Generally speaking, economists tend to avoid nominal values since real values tell us so much more. The term 'nominal' means 'face value' or 'money sticker price', while 'real' is a way of comparing the nominal (face) value with a given **base value** to see what the actual change is. Putting nominal values into real terms is also known as using *constant prices*, i.e. output valued at a price level of a given – base – year.

If output in 2000 is 100 Widgets at a value of €50 then *nominal* output is €5,000. Now, if we call the year 2000 our **base year** (or period) then this is the period we will refer to when comparing all other values; this means that the base year nominal value is also a real value, since all coming output values will be put in terms of base year prices. So, if 110 Widgets are produced in the next time period at a value of €55, nominal output would be €6,050, which is an increase of 21%. Yet clearly real output has only increased by an additional 10 Widgets, i.e. 10%. What we must do is *deflate* the nominal value by removing the inflationary element in order to show real output in the second time period.

Definition: 'Real and nominal national income'ßΣ

Nominal national income is expressed in the current prices of the output period and thus contains an element of inflation. **Real national income** is the nominal value put into base year (or constant) prices to allow real comparisons of output over time. The formula for deflating nominal GDP is:

$$GDP_{real} = \frac{GDP_{nom} \text{ of year measured}}{\text{Price index of year measured}} \times 100$$

While the consumer price index (CPI) is the most commonly referred-to measure of inflation, it is too narrow for general use in macroeconomic models. This is alleviated by using another index of the price level, the **GDP deflator**, which includes **all goods accounted for in GDP**. Since the CPI only includes consumer goods, a great deal of important prices are left out, notably those for investment goods, government goods and exported goods. The GDP deflator is calculated using the same basic principles as in the CPI, but includes prices on all these types of goods in the economy rather than just consumption goods. It is this wider definition of the price level we use to arrive at real GDP and real GNP. So, let us assume that in factoring in price changes for I, G and X we get the following GDP deflator series:

MACROECONOMICS

Figure: 37.5 The GDP deflator

Dec 1995 (t_0)	**Dec 1996 (t_1)**	**Dec 2001 (t_2)**
CPI (t_0) = 100	CPI (t_1) = 102	CPI (t_2) = 119
... plus I, G and X goods	... plus I, G and X goods	... plus I, G and X goods
= GDP deflator 1995: 100	**= GDP deflator 1996: 102**	**= GDP deflator 2001: 128.5**

The GDP deflator shows that the average price level of all goods and services in the Irish economy has increased by almost 30% during the period 1995 to 2001, far more than the 19% shown by the CPI. Keep in mind that our 'basket' is a representation of the **average** change in price rather than a change in **all** prices!

Polishing off the example above, Ireland had the following GDP figures in current (i.e. nominal) values; 1995 = €52,641 bn; 1996 = €58,080 bn; and 2001 = €114,744 bn. Notice that national income looks to have *more than doubled* in a six year period! However, by using the GDP deflator series in Figure 37.5 to deflate these figures, we remove the 'pumped up' component, which is to say the inflation element included in the nominal figures. (See Figure 37.7).

Deflating nominal GDP is done by using essentially the same formula as before, but using the *GDP deflator* as the general price index rather than the CPI:

Figure: 37.6 GDP deflator calculation

$$GDP_{nom} = \frac{GDP_{nom} \text{ of the year}}{GDP \text{ deflator}} \times 100$$

Nominal GDP:	year 1995 = €52,641 m	1996 = €58,090 m	2001 = €114,744 m
GDP deflator: Real GDP:	$\frac{100}{\text{year 1995} = €52,641 \text{ m}}$	$\frac{102}{1996 = €58,090 \text{ m}}$	$\frac{128.5}{2001 = €114,744 \text{ m}}$

Adjusting the current (nominal) values for the increase in general price level

$$\left[\frac{GDP_{nom} \text{ at year measured}}{CPI \text{ at year measured}} \times 100 \right]$$

we get the following real GDP figures illustrated in Figure 37.7 below. This is the 'shrinking balloon effect' of deflating nominal values into real values.[6]

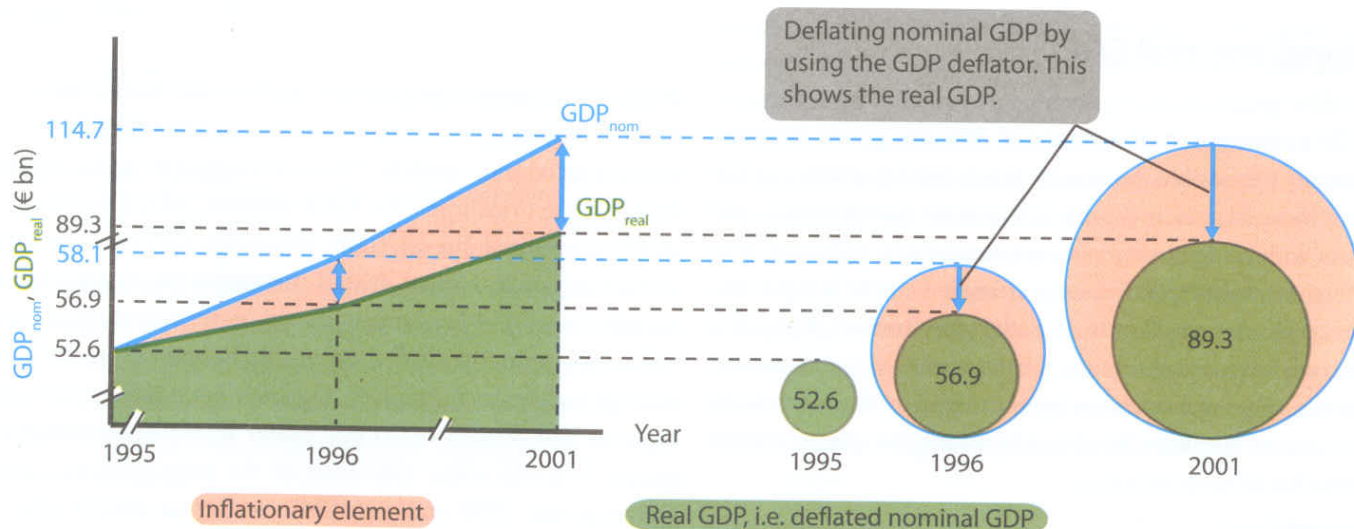---

6   I have rounded the figures a bit.

Figure 37.7 'Pricking the balloon' – deflating nominal GDP in Ireland 1995 – 2001

After adjusting the nominal figures for inflation and thus putting them into base year values, we see how a portion of nominal GDP was in fact comprised of increasing prices rather than increasing output. After deflating these values we get a GDP series showing output in constant year 1995 prices. This is real GDP in terms of base year values of output.

Figure 37.8 Nominal and real GDP in Ireland, 2001 (billions of € at constant 1995 values)

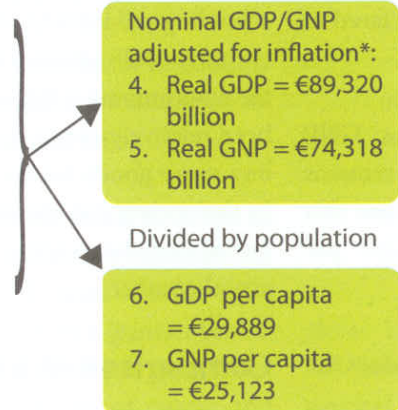| 1	Nominal GDP	114,744
2	Deflated using GDP deflator (index)	128.5
3	Equals: Real GDP	89,320

Source: *National Income and Expenditure 2002, Irish Central Statistics Office,* http://www.cso.ie

In summary; *real* Irish national income increased by almost 70% during the period 1996 – 2001, which is less than the over 100% which *nominal* income suggests, but still one of the highest growth rates in the industrialised world during this period. Just imagine that you were an 'average' Irishman and that your personal income kept pace with the real GDP – you would be able to buy 70% more goods with your income. And since price indexes have a marked tendency to understate real output by not taking into account quality increases and availability of substitutes, real GDP in all probability increased by more than 70%.

Per capita income

National income is also often put into terms of 'average income per person', which is nothing other than **GDP divided by the population**. This is *income per capita* (= GDP or GNP per head). Irish nominal GDP in 2001 was €114,744, and dividing by the population gives us € 29,889; this is Irish GDP per capita. Figure 37.9 below follows the Irish figures used throughout this section.

1. Gross domestic product (GDP) = €14,744 billion plus net property income from abroad –€8,295
2. = Gross national product (GNP) = €96,448 billion minus depreciation –€11,619
3. = Net national product (NNP) = €84,829 billion

Nominal GDP/GNP adjusted for inflation*:
4. Real GDP = €89,320 billion
5. Real GNP = €74,318 billion

Divided by population

6. GDP per capita = €29,889
7. GNP per capita = €25,123

Divided by population:
8. Real GDP per capita = €23,266
9. Real GNP per capita = €19,359

*Base year is 1995 – so all real values are in constant 1995 euro

Figure 37.9 Summary of national income accounting – Ireland in 2001

I offer you a piece of advice here: **Beware of averages!** Bill Gates did a world tour in 2003 of developing countries in furtherance of the Gates Foundation. Find a picture of Gates sitting with a few hundred villagers and then calculate the average income of the group! See my point? The lesson is simple; the more divergence between the maximum and minimum values, the weaker the arithmetic average is. Whether looking at real or nominal GDP or GNP per capita it is essential to keep in mind that *income distribution* can be such that a small proportion of the population accounts for most of the income. This doesn't show in per capita GDP/GNP figures.

MACROECONOMICS

POP QUIZ 37.1

1. Why might an LDC have higher GDP than GNP? Why might Singapore?

2. What is the sum of **global** net property income from abroad?

3. A Volkswagen Polo cost SEK (Swedish Crowns) 65 000 in 1985 and SEK113 000 in 2000. CPI in 1985 was 154 and 258 in January 2000. What is the **real price** of the Polo in 1985 and 2000 respectively using 1980 as the base year? Any weakness in comparing prices like this?

4. In a country, you have the following figures: GDP = 600; property inome paid abroad = 120; net property income from abroad = 30; capital consumption = 60. What are the values of GNP and NNY?

5. Let us say that in 1987 the CPI is 220 and the same year nominal GDP is €600 bn. In 1999 the CPI is 330 and nominal GDP is €900 bn. How has this country fared in terms of real output (= real GDP).

6. Over a 10 year period, the following happens; GNP (nominal) increases by 25%, population remains unchanged, CPI goes from 150 to 200. How has this affected nominal GNP per capita and real GNP per capita?

7. Refer to the figures below. During which period does the rate of inflation fall **first**?

	1972	1973	1974	1975	1976
Prices	97	105	106	105	103
Wages	100	105	110	112	109

8. Explain how 'double counting' can occur in calculating national income, and how measuring 'value-added' can overcome this problem.

9. GDP figures for a country are as follows: USD 200bn in 1989; USD 230bn in 1990; and 260bn in 1991. What are the growth rates? Are they rising, falling or constant during this time period?

10. Over a period of time, the GDP index in an economy goes from 100 to 120 while the population index goes from 100 to 130. What could one deduce from this in terms of average national income? What is missing in order to be able to comment on the change in **real** income?

Use and weaknesses of national income figures

> 'I have observed that we all get the same amount of ice. The rich get it in the summertime and the poor get it in the winter.' Bat Masterson[7]

It is extremely important to understand the uses of growth and GDP figures and also the limitations. The figures are useful for assessing how well the economy is using resources and how the economic system compares to other countries'. The figures are also invaluable in showing possible government policies and future government tax revenue and thus how well pensions, schooling and infrastructure will be covered. Having said this, it is vital to realise that GDP and growth figures do not show how the environment is impacted, whether the increased wealth has been relatively evenly distributed, whether people can actually buy more goods for their money or whether general welfare in the society has increased correspondingly. Taken together, one should always try to use a critical eye and 'look behind' the actual numbers.

Comparison in a single country over time

There are a number of criticisms levelled against GDP figures when they are used to show growth **within** a country over time. Three main points emerge:

1. **Money values and population change:** We have shown how changes in price levels and population can skew income figures and how this can be dealt with by putting all figures in constant and/or per capita values. However, there is still quite often a…

2. **Quality and substitution bias:** Which goods are included in the basket which is compared over time? Price indexes frequently *overestimate inflation* and thus underestimate real output growth over time due to the fact that when aggregate demand changes permanently over time there has been a shift to other, superior,

7   Legendary gunfighter in the 'Old American West'.

goods; *substitutes*. This causes an overestimation of the price level since a downshift in production of goods on the way to becoming obsolete (= outdated) reduces the possibility of benefits of scale and keeps costs and prices higher than for new goods not yet included in the basket. Similarly, GDP figures cannot estimate the *increased quality* of new goods, for example the fact that newer cars need less service and maintenance and use less fuel. Thus, there is a bias in overestimating actual price increases by underestimating the quality of new goods substituting old goods.

3. **Errors and/or changes in accounting methods:** Imagine the hundreds of millions of figures entering into the overall calculation of national income during a year – there are bound to be both *accounting errors* and *time lags* in assessing all the data. It is also often the case that older data is less comparable with newer, due to different methods of accounting.

Comparing GDP between countries

In addition to the above, a number of weaknesses in national income figures become apparent when different countries are compared. These include (but are not limited to!):

1. **Composition of output:** Perhaps the main criticism of GDP as a measure of welfare is that however accurate the figures are for output, the final GDP figure does not show what is being produced – we're back to a 'guns or butter' problem. The Soviet Union of the 1930s put great effort into competing with the west in terms of output and growth in order to show the superiority of the centrally planned system. Going by official output figures the USSR won the race, but what the figures do not show is that the Soviets put the majority of resources into producing capital goods – and never really got around to providing for the wants and needs of its citizens in terms of consumer goods. A country with double digit growth rates and empty shelves is something of an anomaly but quite possible, which is also often the case in times of major conflicts when armaments account for economic growth which in no way represents an increase in the standard of living.

2. **Composition of expenditure:** In a vein similar to the above, national income figures are often skewed by the simple fact that different countries will have different expenditure patterns. The cold and icy Nordic countries

spend a sizable proportion of their income on heating homes and offices – but this doesn't mean a higher standard of living than in temperate climates. Comparing Finland's GDP figures with Bermuda's could get tricky indeed.

3. **Distribution of income:** All per capita national income figures are **averages** and therefore neglect how income is distributed amongst citizens. The third richest country in the world, the USA, had a per capita GDP of USD35,200 in 2000 while the second richest, Norway, had USD37,200.[8] Yet the highest 10% of all income earners in America accounted for 30.5% of national income – and at the same time the country had some 30 million people living below the official poverty line.[9] In Norway, on the other hand, 21.8% of income went to the richest 10% – thereby accounting for one third less than the USA's richest upper deciles.

4. **Unaccounted-for activity:** Statistical inaccuracies will be enhanced by the fact that a portion of economic activity will be hidden; *parallel markets* for goods and labour are notable examples of a type of *systematic error* since real output is consistently under-reported. The large element of barter and non-money economic activity in LDCs will lead to consistent under-reporting of real output figures. More developed countries will also have a large section of unreported activity, but in this case primarily due to tax avoidance and evasion of labour laws.

Note: Large parallel markets can seriously underestimate economic activity. For example, unreported national income on a global basis amounts to almost 31% of the world's GDP.[10] The range is from 9% of GDP in the US to over 70% in Bolivia and Georgia. Indonesia is estimated by the OECD to have a parallel market of close to 80% of all economic activity. Such figures seriously undermine the validity of national income accounting figures.

5. **Exchange rates distortions:** Since the comparison of different countries' GDP must be put into some form of common language, a single currency is used, often the USD. In doing this, the market exchange rates for

8 National Accounts of OECD countries, Main aggregates, Volume 1, updated version from 2003

9 *Poverty in the United States*, US Census Bureau 2000, page 3

10 Schneider, Buehn, Montenegro, "Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007", World Bank Discussion Paper at http://www.econ.jku.at/members/Schneider/files/publications/LatestResearch2010/ShadEcWorld10_2010.pdf

MACROECONOMICS

different currencies are used – with the unfortunate side effect of grossly underestimating average incomes of low cost countries when *purchasing power* is taken into account. See purchasing power parity (PPP) below.

6. **Externalities and environmental damage:** GDP figures do not show soil erosion, air pollution, land degradation, deforestation, depleted natural reserves of resources or the often monumental disruptions to values and traditions as a result of large scale economic growth in a relatively brief period of time.

In spite of the heavy criticism levied above, GDP per capita adjusted for purchasing power (see below) is still the best 'single' indicator of development available. There is relatively clear positive correlation between most standard of living indicators and economic growth – the problem being that it is very difficult to see which comes first, i.e. there is a causality problem.

Green GDP

There have been several efforts to factor in the external costs of production into country wide figures of output; "green GDP accounting". The latest effort was done in China during the period 2005 to 2007 – the hitherto only effort at green national income accounting. No official figures were published. If it ever takes off I shall include it here. Watch this space.

Summary & revision

1. **GDP is accounted** for via three methods; the **output method** (sum of value-added in an economy); the **income method** (sum of wages, rents, interest and profits); and the **expenditure method** (sum of consumption expenditure, investment, government spending, and net exports).

2. **GNP** is GDP plus *net property income* from abroad.

3. **'Domestic'** means '…produced within a country's borders.' and disregards who owns the factors. **'National'** looks instead at '…who owns the factors – regardless of where production takes place…'

4. **Real GDP** (or GDP at constant base year values) is nominal GDP adjusted for inflation. To calculate real GDP, divide nominal GDP by the GDP deflator (a broad price index) and then multiply by 100.

$$GDP_{real} = \frac{GDP_{nom} \text{ of year measured}}{\text{Price index of year measured}} \times 100$$

5. **Per capita GDP** or GNP is calculated by dividing GDP or GNP by the population.

6. National income figures are used to compare the overall 'success' of an economy with others and to provide statistical material for governments to base future decisions on.

7. There are several **weaknesses** of national income accounts:

   • **Changes in population** and money values skew the figures

   • **New and better products** are not represented in the money-value-of-output figures in GDPGNP

   • Simple **calculating errors** are all too common

   • *What* is produced is not shown in the figures – e.g. **composition of output** is not reflected in a GDP value

   • National income per capita is an *average* and does not show income distribution

   • Many economies have large unaccounted-for sectors in the economy – so called **parallel markets**

   • **Exchange rates** tend to distort comparisons between economies

   • **Environmental impact** of production is notoriously missing (and difficult to estimate!) from GDP/GNP figures

38. Calculating GDP and GNP in Nominal and Real Values

Key concepts: HL extensions

- Calculating nominal GDP
- Calculating GNP/GNI
- Calculating GDP using the GDP deflator

The tables below show figures from the Irish National Accounts and are based on the expenditure method of national income accounting.[1] Some figures have been left out and you don't need a degree in rocket surgery to understand your task. Grab a piece of paper guys and fill in the blanks. Answers in the "Summary and revision" section.

Calculating nominal GDP

Nominal GDP is the sum of the money value of all expenditure of an economy during a period of time – here 2008 to 2010 in Ireland.

Item (millions €) at current market prices	2008	2009	2010
Consumption expenditure (goods and services)	92,098	81,779	79,292
Government expenditure	33,528	31,938	29,522
Investment expenditure (fixed capital formation)	39,430	25,293	18,074
Exports of goods and services	150,181	145,902	157,673
Imports of goods and services	-133,877	-121,037	-127,901
Value of physical change in stocks	-596	-2,264	-852
Statistical discrepancy	-774	-1,015	184
GDP at current market prices			

1 http://www.cso.ie/en/statistics/

Calculating GNP/GNI

GDP looks at 'where goods are produced' rather than 'who owns the factors of production'. In moving from GDP to GNP we are accounting for ownership of factors abroad. Ireland will repatriate (bring home) profits from abroad (factor income from abroad) and foreign firms with holdings in Ireland will send profits abroad (factor income sent abroad). Revise the term 'net' and fill in the blanks for GNP at market prices. Then revise 'depreciation' (or 'capital consumption') and fill in GNI at market prices.

Item (millions €) at current market prices	2008	2009	2010
GDP at market prices (e.g. nominal GDP)	179,990	160,596	155,992
Net factor income from world	-25,317	-28,363	-27,785
GNP at market prices			

Calculating real GDP using the GDP deflator

I am most insistent with my students that they understand on an almost intuitive level that '…nominal means nothing and *real* means everything…' If the money value of GDP increases by 5% and prices have increased by 5%…is the economy in fact producing more 'things'? No, of course not. Since we are measuring the money value of output we must remove the inflationary element to see whether in fact more 'things' are being produced. Nominal income values are adjusted for inflation by using the GDP deflator:

$$GDP_{real} = \frac{GDP_{nom} \text{ deflator of year}}{GDP \text{ deflator of year}} \times 100$$

Item (millions €) at current market prices	2008	2009	2010
GDP deflator (base year is 2008)	100	95.9	93.5
GDP at constant 2008 prices	179,990	167,462	166,836
GNP at constant 2008 prices	154,673	137,886	137,120

Now, because I am just as irritating as all your other teachers, here are a few questions based on the data above:

POP QUIZ 38.1

1. How has the Irish economy fared during the period 2008 to 2010? What has the growth rate been? Think first – remember that you are an *economist*!

2. Your answer to 1. above should come with '…we actually *also* need to know…'. What else do we need to know?

3. What is the main component of expenditure in the Irish GDP (in percentage)? Calculate.

4. Having been a diligent student and looked things up, you can now address the question of why GNP is in fact *smaller* than GDP. (OK, admittedly you can simply revise Chapter 37.)

5. What has happened to the price level in Ireland between 2008 and 2010? Look up the correct definition of this situation.

6. After at first falling, export earnings rose between 2008 and 2010. Look at your answer to 4. above and speculate as to why there might be strong correlation (and causality!) between exports and domestic price level.

Summary & revision

1. The Irish economy was hard hit by the global recession. In nominal terms GDP shrank by 10.78% in the period 2008 to 2009. GDP fell again in 2009 to 1010 by 2.86%.

   a. However; these are nominal values and unadjusted for inflation! Since there was negative inflation (a fall in the general price level) we know that in real terms the decrease in GDP will not be as bad.

   b. Using real GDP figures for the two periods, we get –6.96 and –0.37; less than the nominal decline in GDP.

2. We would need to know whether the population of Ireland changed significantly during the period. Many Irish have families and links to the US and Australia and many foreign workers live in Ireland – net emigration due to 'bad times' might lead to a decrease in the population.

3. Main component is in fact consumption expenditure which accounted for just over 50% of GDP in nominal terms 2010. In contrast, net exports [export revenues minus import spending] comprised 19% of GDP.

4. Ireland has attracted huge amounts of foreign direct investment (FDI) and therefore will have larger outflows of property income than inflows. This gives a net minus value of net property income flows so GNP will be lower than GDP.

5. The price level – as measured by the GDP deflator rather than the CPI – has fallen during the period. This is known as deflation.

6. Assuming *ceteris paribus*, a fall in Ireland's general price level means that Irish exports are cheaper relative to other countries. A fall in export prices would increase demand for Irish goods and so increase export revenues. However, this is in no way certain as it depends on the price elasticity of demand for exports. (See Chapter 73 for depth on this issue.)

39. The Business Cycle

Key concepts:

- Short term fluctuations and long term trend – phases of the cycle
- Decrease in growth and decrease in growth rate

"Recession is when your neighbour loses his job. Depression is when you lose yours. And recovery is when Jimmy Carter loses his."
(Ronald Reagan during 1980 US presidential campaign)

Short term fluctuations and long term trend – phases of the cycle

There should be no difficulties for you in finding your way through Figure 39.1. The long run trend (potential real GDP) is upward sloping and in keeping with empirical evidence which suggests a long run trend of around 1.5 to 3.5% yearly growth. (Yet note that the cycles are highly stylised.) The cycle is discernible as reoccurring expansions and contractions. Economic activity – measured by real GDP – at its lowest point is called a 'trough', followed by 'recovery', 'boom' and 'peak'. When economic activity slows and then falls over a period of time, one speaks of 'recession'. When the bottom of the cycle is reached once more a cycle has been completed – which is also measured as the time periods from peak to peak.
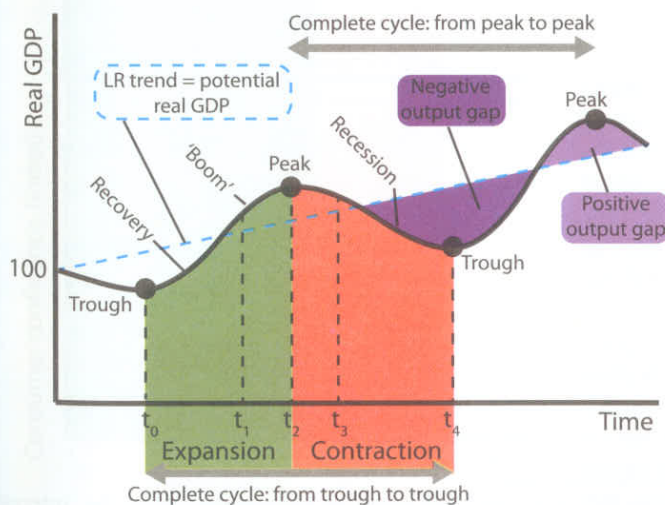
The cycle in Figure 39.1 also shows the difference between potential and actual output in the **output gaps**. The output gap is defined as actual real GDP minus potential GDP, and one therefore speaks of *negative output gaps* when de facto real GDP is lower than potential and *positive output gaps* when real GDP is higher than long run potential. Over the cycle, the components of aggregate demand and variables within the main macro objectives will be affected in many ways. Here is a general outline:

Trough: This phase is marked by overall low economic activity (compared with earlier periods) and increased unemployment, as firms lower output and workers are made redundant (= laid off). Lower incomes will decrease consumption expenditure in the economy and there will be low (-er) inflationary pressure. In general, the period is one of low consumption, low investment and also low imports which will be affected by the fall in both consumption and investment. Lower government tax revenues coupled with high transfer payments (unemployment and social benefits for example) often lead to budget deficits for governments (see also Chapter 57 and 'automatic stabilisers').

Recovery: When the economy picks up, i.e. output increases due to demand and/or supply variables, GDP increases, unemployment falls and consumption, investment and imports rise. The period is commonly associated with an increase in **inflation** and rising interest rates.

Boom/peak: When growth rates increase over shorter time periods, the economy can 'overheat'. Firms are nearing capacity, labour markets are tight (i.e. labour is becoming scarce due to output levels beyond the full employment level) and wage levels are increasing and so feeding inflation. Consumption, investment and imports are increasing – as are tax revenues. There are often speculative bubbles during this phase, as high inflation induces increased demand and subsequent speculation for items considered to be good value-retainers; land, property and shares are examples herein.

Figure 39.1 Business cycle

MACROECONOMICS

Recession: When, for any number of reasons, economic activity abates (= grows less) to the point where real GDP actually falls, one speaks of **recession**. Firms will lower investment and output while seeking to rid themselves of labour no longer needed. Households will be affected by rising unemployment and decrease consumption and thereby also imports. Government tax revenues fall and transfer payments rise. Inflation rates decrease as do interest rates. (In fact there might be **deflation**; falling prices.)

Definitions: 'Inflation, deflation, and recession'

Inflation is defined as a sustained general increase in the average price level and measured by the change in CPI.

Deflation is a sustained general decrease in the average price level.

Recession is when an economy experiences two consecutive quarters (six months) of falling real GDP. (This can in fact not be seen in figure 39.1 since there is no actual time line showing the months.)

It is worthy of noting that the long term trend is by no means 'set in stone' – in other words, it too can change. The US economy showed a long term growth rate of 2.44% between 1928 and 1972, slowing to 1.93% from 1972 to 2007.[1]

Decrease in growth and decrease in growth rate

One common mistake seen in many an exam paper is when students fail to clearly distinguish between decreasing growth (e.g. negative growth) and a decrease in the *rate* of growth. In the figure 39.1, the period just before the 'peak' is characterised by a slower *rate* of growth – the slope of the business cycle curve falls. The period from t_1 to t_2 shows a slower rate of growth while t_2 to t_3 shows a decrease in growth. During the period 2006 to 2007, the US growth rate (in GDP terms) fell from 2.7% to 1.9% - a fall in the *rate* of growth. During the economic crisis of 2008 and 2009, US growth went from 0% in 2008 to a low of –3.5% in 2009 – e.g. the US economy *contracted* by 3.5% during 2009.[2]

Causes of cyclical variations seen in the business cycle

Perhaps the main question in business cycle economics is what **causes** the changes in economic activity. The simple answer is that there is no single dominant theory able to explain the fluctuations fully. A rather unsatisfactory answer is that when aggregate demand or aggregate supply change, then GDP is affected and so too is the business cycle. A far more complex – and imprecise – answer deals with the array of *underlying variables* which in turn influence the components of aggregate demand and the influences on aggregate supply. We will cover some of the main cyclical variables in using the AS-AD model in Chapters 41 and 42.

A question I invariably get at this stage shows not only insight but an intuitive feel for economic relationships; how do we know what the potential real GDP trend line looks like over longer periods of time? The answer is that it is impossible to set a precise value for the long run trend of output in the economy. A trend over time can be calculated by adapting a trend line which removes cyclical variations to a series of GDP values over time. Without beating you over the head with mathematical computations – that is best left to your maths teacher – the method basically takes real GDP values and plots them over a given time period. Then a line of best fit is adapted to give the long run trend line.[3] One of the main goals in business cycle research is identifying key turning points, i.e. the troughs and peaks, over the longer term, in order to adjust monetary and fiscal policies correctly. This has met with limited success so far and economics is still far better at predicting where we've *been* (on the business cycle) than where we *are* – or indeed, where we are *heading*.

1   *Why one economist predicts slow growth*; Bisiness Week, 30th September 2010

2   www.data.worldbank.org/indicator

3   In fact, the trend will often be exponential since the 'interest upon interest' effect exists. The trend line is therefore in most cases a log value against the Y-axis. Pester your math teacher for details. Don't give him/her my email address.

MACROECONOMICS

A Little Depth

Figure 39.2 Predicting cyclical variations in the business cycle
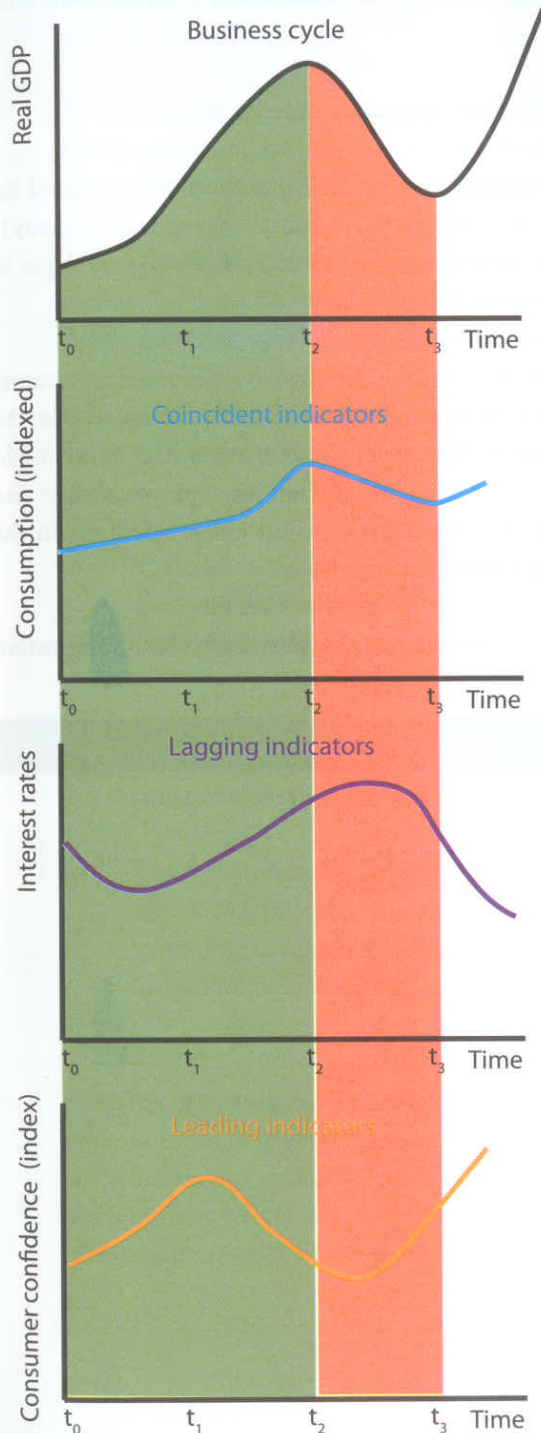
Business cycle

A key element in studying the cyclical variations in economic activity is finding variables that show correlation with the business cycle. To this aim, a number of indicators have been noticed to 'shadow' changes in real GDP. For example, unemployment falls during a recovery (called counter-cyclical variables) while consumption increases during the same period (pro-cyclical). Three main groups of indicators have been identified showing to coincide, lag and precede changes in real output.

Coincident indicators

A coincident variable is broadly in line – in terms of a time line – with the business cycle. Industrial production, investment, consumption and imports are all examples of pro-cyclical coincident indicators.

Lagging indicators

Inflation and interest rates are pro-cyclical and lagging, as are housing prices and number of bankruptcies. Unemployment is counter-cyclical and rather markedly lagging, since employers do not want to lay-off workers and delay this as long as possible. Thus unemployment levels can lag output changes by as long as 12 to 18 months.

Leading indicators

Perhaps the most closely watched of all indicators of economic activity are those that seem to precede changes in real output, since such variables would have strong predictive value. Key indicators (all pro-cyclical) are inventory levels (as firms build up stocks in anticipation of increased demand), stock market index, new car sales, consumer confidence (survey-based), initial claims for unemployment benefits, price of raw materials and new orders levels to firms. These are often weighted, indexed and used to construct a composite index of leading indicators which is closely followed by firms and policy makers alike. (Note the element of 'nonsense correlation' (see Section 1) herein. Increased sales of new cars do not cause GDP to increase, but serve instead as an indicator of far more complex undercurrents in the economy, say 'warm fuzzy feelings' in households causing people to have confidence in the economy and subsequently go for the 'big ticket items' such as new cars.

MACROECONOMICS

273

The importance of studying the business cycle

Studying business cycles has great importance for the improvement of economic policy. Extreme fluctuations are very damaging to the economy for several reasons:

- First and foremost, swings in the economy lead to great *unpredictability* for firms and households, making it difficult for firms to plan both input and output and adding insecurity to households.

- Cyclical variations can also create excessive swings in business activity and heighten *social costs*, primarily by inducing firms to increase investment during upturns, which builds in redundant capacity during downturns, which increases unemployment.

- Governments are also affected by any and all *unplanned changes in tax revenues* and social spending needs, and since both are built-in to the government budget a volatile business cycle adds an unneeded degree of uncertainty in setting a budget for the coming fiscal year.

Many economists would therefore agree that the most desirable macro state in the long run is a *steady increase* in real GDP with minimal fluctuations – but there is wide disagreement as to both the causes and solutions of business cycle fluctuations. As the next section will show, there is a great divide between proponents of Keynesian and neo-classical theory on the subject of economic policies aimed at evening-out the business cycle while enhancing growth in the long run.

POP QUIZ

Business Cycles

1. How would the main macro objectives be affected over different phases of the business cycle?

2. Depict all four stages of the business cycle using the AD/AS model.

3. Explain why an output gap can be inflationary.

4. Here's a tricky one. Assume that the labour demand for nurses is not directly linked to changes in economic activity. Why might nurses still receive higher wages in an economic boom period?

5. Say that firms in an economy have increased investment a great deal during an 'overheated' period and that the economy suddenly moves into recession. How might this increase in capital worsen unemployment and also delay a movement back to full employment when the economy starts to recover?

6. Why is understanding the business cycle so important for economic policy makers?

MACROECONOMICS

Summary & revision

1. The **business cycle** roughly follows the phases of trough, recovery, boom, peak, recession.

2. **Recession** is defined as falling real GDP over two consecutive quarters.

3. The **long run trend** is calculated as the long term average growth rate.

4. A **decrease in growth** (measured by a change in real GDP) means that national income has decreased during a period of time, usually over the course of a year.

5. A **decrease in the rate of growth** (change in real GDP) means that national income is still increasing but at a slower rate over a period of time, usually one year.

6. Economists have identified several **indicators of economic activity** that are correlated to the business cycle.

    a. *Coincident indicators*; changes that are broadly in line with the business cycle (industrial production and consumption for example)

    b. *Lagging indicators*; changes in economic activity occurring with a time lag such as interest rates and unemployment

    c. *Leading indicators*; changes in the economy which occur before cyclical swings – new car sales, consumer confidence and raw material prices are commonly watched indicators

7. **Studying the business cycle** is important for several reasons. It can help *governments plan policies* of taxation and government spending and the *lack of predictability* has negative effects on household consumption and investment expenditure by firms.

MACROECONOMICS

40. Aggregate Demand

2.2

Key concepts:

- Simple demand vs aggregate demand (AD)
- Components of AD
- Why the AD curve is downward sloping

'It's the economy, stupid!'
Campaign slogan for Bill Clinton dreamt up by his advisor James Carville.

Simple demand vs aggregate demand (AD)

Let us point out a few basic similarities to the simple demand curve in order for you to have something to hook your learning curve on to – all the while keeping in mind that aggregate supply and demand models are far more complex since we are dealing with all goods and services rather than a single good.

1. Just as in simple demand (demand for a single product), aggregate demand curve is downward-sloping, which shows *negative correlation* between the price level (measured by the GDP deflator) and real output (measured by real GDP). This is quite similar to the simple supply and demand model. However, note that it is impossible to use 'quantity' on the X-axis when dealing with the aggregate since there is a slight difference between paper clips, nuclear submarines and deep sea oil rigs. To overcome this, we use the inflation-adjusted *monetary* value of aggregate output, e.g. *real* GDP.

2. The pattern of correlation shows that a fall in the average price level (GDP deflator) will lead to a movement along the aggregate demand curve – i.e. an increase in real GDP at a given price level. Didn't we just forget something?

3. Yes, that all other influences on aggregate demand are assumed to be held constant, i.e. *ceteris paribus*. Once again our model is based on the concept that the pattern of correlation will hold true only if all other variables are left unchanged. (See *Shifting the aggregate demand curve* following.)

Components of AD

Aggregate demand is the sum of total planned expenditure for a given price level by households, firms, government and foreign sectors during a period of time. The components of aggregate demand have been covered in Chapter 37 in the **expenditure method** of GDP accounting:

- **Consumption expenditure** (C in the formula below); households' spending on goods and services

- **Investment expenditure** (I); firms' expenditure on capital goods such as machines and factories – this is in addition to the capital stock of the economy (do NOT confuse investment with "…buying shares…" – this is in fact a form of saving)

o *Gross* investment is when firms increase spending on capital goods due to an increase in demand for their goods

o *Net* investment occurs when firms replace worn-out capital (also known as replacement investment)

- **Government spending** (G); government spending goes to health care, defence, roads etc. It is assumed to be unlinked ('exogenous' in econ-speak) to the price level, e.g. determined by political decisions at local, regional and national level.

- **Export revenue** (X); spending on domestic goods by foreign firms, households and governments

- **Import expenditure** (–M); expenditure on foreign goods by domestic firms, households and firms (note that this component must be deducted since foreign goods and services are not part of GDP)

Aggregate demand is the sum of expenditure components, which gives us:

$$AD = C + I + G + X - M$$

It is worthwhile noting at this early stage the italicised 'planned'. In the previous circular flow model all expenditure was de facto (= actual, realised) GDP, since all output also represented expenditure and income – it would be impossible for expenditure to exceed output. In the aggregate demand/ supply model we attempt to show a more dynamic and realistic view of the macroeconomic world by showing what happens when planned expenditure increases at a different pace than de facto output, e.g. inflation is brought into the picture.

The aggregate demand in an economy is dependent on a great many influences, which is to say that the individual components of aggregate demand are influenced by a great array of underlying variables. And similar to the previous case of (simple) supply and demand, a model is built to show the relationship between price and quantity – yet since we are dealing with all goods and all prices we have to expand the model a bit.

Why the AD curve is downward sloping

Figure 40.1 shows the relationship between aggregate demand and the average price level. The AD curve shows the connection between the price level and real GDP. The GDP deflator puts the

price level into constant (real) prices (here, 1995) and the four points, A, B, C and D, show how the quantity demanded of real GDP increases as the price level falls. If we were still dealing with simple demand, say video films or bicycles, we would explain this by saying that there is an *income effect* (lower prices mean that people have more spending power) and a *substitution effect* (people change consumption habits and switch to other goods). Both of these effects are in fact present in the downward slope of the aggregate demand curve but there is a greater level of complexity when dealing with the aggregate.
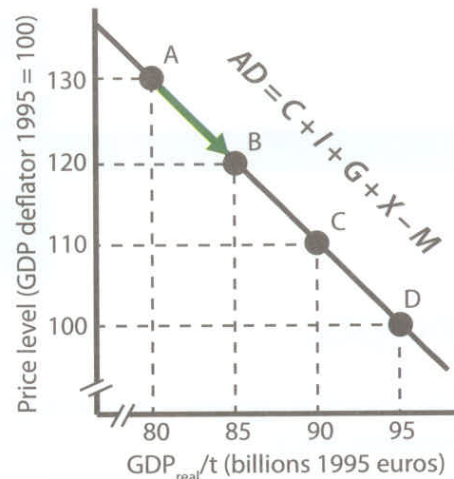

Fig 40.1 The basic aggregate demand model

The AD curve shows the relationship between the price level (measured by the GDP deflator) and real GDP (constant values). The AD curve is downward sloping for three reasons:

1. **Real income effect**; when the price level falls, real income rises (wealth effect)

2. **Real balance effect**; a fall in the price level means that the opportunity cost of saving increases, so households and firms lower saving levels and increase expenditure (Interest rate effect)

3. **International substitution effect**; lower prices - *ceteris paribus* - will increase exports and decrease imports (Exchange rate effect)

Why the aggregate demand curve slopes downward

It is most important to keep in mind that in drawing the curve we are assuming *ceteris paribus*, so only the price level changes; other macro variables such as interest rates, income for trade partners, domestic incomes, inflation…etc *remain the same*. The income and substitution effects from the simple demand model are discernable in the three main reasons for a downward sloping aggregate demand curve:

1. **Real income effect**: A fall in the general price level means that real income has increased, assuming, as we are, that all else remains the same. In this case, we are assuming that the amount of money flowing around the system – the nominal quantity of money has not changed and neither have wages – therefore people will increase their planned expenditure and quantity demanded of real GDP will rise. Referring to Figure 40.1, if the price level goes from 130 to 120, the quantity demanded of real output will increase from €80 billion to €85 billion; this would move us along the aggregate demand curve from **point A** to **point B**.

    The aggregate demand/supply model is based on real output which means that GDP is based on *real* values of money in terms of purchasing power. A fall in the price level means that the purchasing power of any given quantity of money increases, resulting in a de facto increase in real incomes. We are in essence positing that the larger the amount of real money in the economy, the higher the quantity of goods and services – real GDP – will be demanded.

    In economic short-hand: $\Delta\downarrow$price level $\rightarrow \Delta\uparrow Y_{real} \rightarrow \Delta\uparrow C \rightarrow \Delta\uparrow Q_{AD}$

2. **Real balance effect**: Another way of looking at the downward slope of aggregate demand is that both households and firms try to strike a *balance* between planned *present* and *future* investment/consumption. Future investment/consumption is to a large extent the savings of the present. When households and firms save they are in fact putting off present expenditure to the future. The balance between present expenditure and future expenditure is disrupted when the price level changes, since lower price levels mean that firms' and households' savings are worth more in *real* terms. (This is analogous (= similar) to the real income effect above.)

A fall in the price level means that firms and households can save less but still retain a satisfactory level of future consumption/investment. With less being saved, consumption and investment expenditure rises, leading to a movement from A to B in Figure 40.1. This is an example of the **inter-temporal substitution effect**. Again, don't be alarmed by the term 'inter-temporal'. It simply means between ('inter-') times ('temporal') and deals with how households and firms might substitute planned future consumption/investment by bringing some of it to the present. Since the opportunity cost of decreasing present saving has gone down – due to the fall in the price level and the increase in real savings – firms and households will be able to increase consumption and investment sooner rather than later.

In economic short-hand: $\Delta\downarrow$price level $\rightarrow \Delta\uparrow$value real savings $\rightarrow \Delta\uparrow$potential *future* C & I $\rightarrow \Delta\downarrow$opportunity cost of *present* C & I $\rightarrow \Delta\uparrow$C & I $\rightarrow \Delta\uparrow Q_{AD}$

3. **International substitution effect**: Two components of domestic aggregate demand deal with the foreign sector, namely exports and imports. Assume an open economy and that the domestic price level falls *relative to trade partners*.[1] Domestic firms and households will substitute some of their expenditure on imported goods for domestic goods; import expenditure will decrease. (Since imports are a minus value in aggregate demand, lower import expenditure will increase aggregate demand.) Conversely, a lower price level – assuming *ceteris paribus*, i.e. the exchange rate and price level in trading countries remain the same – will lead to increased quantity demanded from our export sector. This means that **net exports** (X-M) increase, leading to an increase in quantity demanded of real output; a movement from **A** to **B** in Figure 40.7.

In economic short-hand:

$\Delta\downarrow$domestic price level $\rightarrow \Delta\uparrow$*relative* $P_M \rightarrow \Delta\downarrow M \rightarrow \Delta\uparrow Q_{AD}$

And:

$\Delta\downarrow$domestic price level $\rightarrow \Delta\downarrow$*relative* $P_X \rightarrow \Delta\uparrow X \rightarrow \Delta\uparrow Q_{AD}$

---

1   Inserting 'relative' here is rather important. If our trade partners' average prices increase by 5% and our domestic prices increase by 3% then domestic prices have fallen *relative* to our trade partners'.

MACROECONOMICS

DANGER MINES **WARNING !**

Macroeconomic models: Confusing a change in price level with a change in interest rates

A strong influence in aggregate demand is the rate of interest, i.e. the 'price' one pays for borrowing money – alternatively, the 'return' one receives for keeping money in the bank. The **interest rate is positively linked to the price level**, i.e. when the price level goes up, so do interest rates. This is because lenders (creditors, e.g. banks) will want to retain the real rate of interest being paid to them by debtors (= borrowers, e.g firms and households), and higher inflation means that banks will have to raise the interest demanded from households and firms in order for future incoming interest flows to the bank to have at least the same value in *real* money terms as when the loan was taken. Another reason is that when the price level rises, firms and households will have to increase their borrowing (= demand for loans increases) in order to retain consumption and investment levels – the interest rises (= price of loans) rises.

So:

$\Delta\uparrow$price level $\rightarrow \Delta\uparrow$r (interest)

and conversely; $\Delta\downarrow$price level $\rightarrow \Delta\downarrow$r

Assume now the economy is at point A on the aggregate demand curve in Figure 40.1 and that the price level falls. This will lead to a fall in interest rates, as explained above.

The lower the interest rate, the lower the opportunity cost of consumption and investment since households and firms forgo less interest on their savings. Households and firms will now be inclined to increase their expenditure. Another way of looking at this is that since the interest rate has fallen, expenditure on items funded by credit – firms' expenditure on capital and households' purchases of durable goods such as cars and furniture – will increase. There is a movement along the aggregate demand curve from point A to point B.

We get:

$\Delta\downarrow$price level $\rightarrow \Delta\downarrow$r $\rightarrow \Delta\downarrow$S $\rightarrow \Delta\uparrow$C & I $\rightarrow \Delta\uparrow$Q$_{AD}$

Now, here's the warning! In the example given, it is the *price level that has changed* – which in turn affected interest rates and thus consumption and investment. This is an increase in quantity demanded of real GDP, i.e. a *movement along* the aggregate demand curve. However, as shall be seen in Section 3.4, an increase or decrease in interest rates will cause, respectively, a decrease or increase in aggregate demand – that is to say, a *shift in the aggregate demand curve*. Whether there is a movement along or shift in aggregate demand depends on which variable has initially changed in the economy; a change in the price level – *ceteris paribus* – leads to an increase in quantity demanded of real GDP. If interest rates change for any other reason, say that the central bank lowers the rates, there will be a shift in the aggregate demand curve.

In summa;

- Δ**price level** $\rightarrow \Delta$r $\rightarrow \Delta$C & I $\rightarrow \Delta$Q$_{AD}$ = *movement along* AD curve
- Δr $\rightarrow \Delta$C & I $\rightarrow \Delta$AD = *shift* of AD curve

POP QUIZ

1. Explain how the variables interest, inflation and consumption are inter-linked. Note that these variables show dual-causality!

2. Which component of aggregate demand is assumed to be independent of the price level? Explain why.

3. Why is it possible for a change in interest rates to be linked to both a movement along and shift of the aggregate demand curve?

4. How does the international substitution effect help create a downward sloping AD curve?

Summary & revision

1. **Simple demand** shows the ability and willingness of consumers to buy a single good. The sum of total demand, the aggregate, shows the demand for all goods and services in an economy.

2. Our **AS-AD model** uses the overall price level instead of 'price' and total output adjusted for inflation – real GDP – instead of 'quantity'.

3. The AD shows the correlation between the overall price level (GDP deflator) and the demand for consumption,

4. The **AD curve is downward-sloping** due to

    a. The **real income effect**; a fall in the price level means that real incomes rise – which leads to increased consumption, investment and net exports

    b. The **real balance effect**; a fall in the price level means that the real value of savings has increased and households can save less and spend more

    c. The **international substitution effect**; a fall in the price level – *ceteris paribus*! – means that the relative price of exports has fallen and the relative price of imports has risen – export revenue increases and import spending decreases
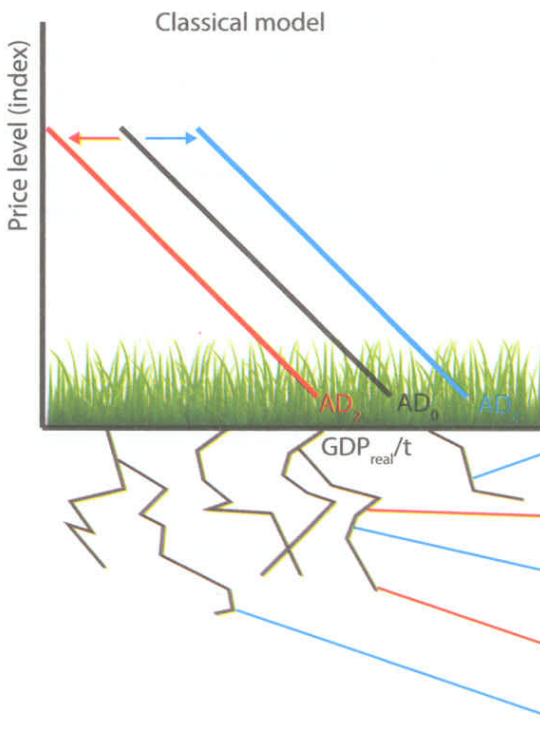
41. Shifts in Aggregate Demand

Key concepts:

- Aggregates supply (AS) and output
- The upward sloping AS curve – short run aggregate supply (SRAS)
- Shifts in SRAS

This brief chapter outlines the various factors which induce (= encourage, cause) changes in aggregate demand, i.e. a shift in the AD curve. Further chapters will go into greater depth on the specific areas dealing with government (fiscal) and central bank (monetary) policies.

Again I become rather childish in resorting to a picture I often use with my students in getting the basic message across on how any change in one of the components of AD will shift the AD curve; I draw grass growing between the components and explain that anything in the grass-roots affecting one or several components causes the AD curve to shift. Figure 41.1 shows the AD curve, grass and root system. The 'root system' is a jumble of hundreds of possible events in the macro environment which can effect the 'grass' – the components of aggregate demand.

A change in any of the components of AD (C, I, G, X or M) will change aggregate demand, i.e. shift the AD curve.

For example:

- Households owning property see housing prices rise and feel wealthier → ↑C → ↑AD
- Firms feel less confident about sales and lower investment levels → ↓I → ↓AD
- Unemployment rises and the government increases spending on infrastructure → ↑G → ↑AD
- Foreign trade partners go into recession and decrease their purchases of Home Economy's goods and services → ↓X → ↓AD
- The Home economy's exchange rate falls making imports more expensive → ↓M → ↑AD

Figure 41.1

Consumption and AD

Changes in household consumption;

Inflationary expectations: Assuming *ceteris paribus*, the expectation of a higher rate of future inflation will lead to increased present consumption. This is based on opportunity cost thinking of households, since buying in the future – when the real value of money has fallen – means foregone consumption possibilities for households. One can also say that

people will choose to hold more of their wealth in real assets (cars, housing, furniture) rather than money assets which are expected to erode ever quicker due to the expected rise in inflation.

- $\Delta{\uparrow}$inflation expectations → $\Delta{\uparrow}$C → $\Delta{\uparrow}$AD ..or..
  $\Delta{\downarrow}$inflation expectations → $\Delta{\downarrow}$C → $\Delta{\downarrow}$AD

Wealth and income expectations: An increase in perceived wealth will increase households' consumption, since an increase in perceived wealth represents an increase in total savings. Please be aware that there is a difference between perceived wealth and de facto (actual) wealth in terms of money. While households holding assets such as property or shares might experience that they are wealthier during a property or stock market boom this is not the same as a de facto increase in monetary wealth, since it is not at all certain that these assets can or will be converted (liquidated) into money at market prices. The effect is instead analogous with the real balance explained earlier; if households feel that future consumption possibilities represented by the perceived value of present assets have risen then present saving can be decreased, leading to an increase in consumption.

Expectations of higher income work in the same direction. If households expect their incomes to rise – perhaps due to successful union negotiations or increased demand on the labour market – then they will react by increasing present consumption. Conversely, every time I almost get fired[1] there's a 'blip' in the McGee planned household expenditure. I am no different from other citizens, in that the prospect of becoming unemployed – or simply reading in the papers about increasing unemployment in the economy – will lead to lower consumption when the image of 'hard times' rears its ugly head.

- $\Delta\uparrow$perceived wealth $\rightarrow \Delta\downarrow S$ & $\Delta\uparrow C \rightarrow \Delta\uparrow AD$ …or… $\Delta\downarrow$perceived wealth $\rightarrow \Delta\uparrow S$ & $\Delta\downarrow C \rightarrow \Delta\downarrow AD$

- $\Delta\uparrow$expected income $\rightarrow \Delta\uparrow C \rightarrow AD\uparrow$ …or… $\Delta\downarrow$expected income $\rightarrow \Delta\downarrow C \rightarrow \Delta\downarrow AD$

Overall positive/negative outlook: I don't know of any mainstream economic term which covers this, so I often simply refer to the 'warm fuzzy feelings' of citizens in an economy. This 'feel good – feel bad' element should not be underestimated! How many times have you not made up for a bad week by going out and spending a bit of money on movies, dinner or shoes? Or the opposite; rewarding yourself after a particularly good week by going to the movies, a restaurant or buying new shoes?! Exactly, your consumption increased in both cases in spite of diametrically opposing feelings. On a macro scale of things, there is a strong element in our consumption habits arising from the basic hopes, fears, joys and sadness of the 'societal organism', as I often call it.

As an example of the above, just consider the attacks on the World Trade Center in the US on September 11th 2001 which caused a decrease in already declining economic activity – both in the US and the world. Insecurity, lack of confidence, fear and overall negative expectations served to lower households' consumption plans and also firms' investment. There was also a strong negative impact on trade, i.e. an *external shock* to aggregate demand.

Household debt: During 2007 the impending economic crisis became increasingly evident to everybody except banks, financial advisers, fund managers, government officials, central bankers, households, journalists…hm, OK, maybe it wasn't that evident.[2] In any case, American households' debt reached the astronomical level of close to 140% of gross disposable income (income after tax) during 2007 while the overheating US economy drove up interest rates. The result was an average household interest payment of 12 to 14% of disposable income. Naturally this had a seriously dampening effect on household consumption; during the first quarter of 2008, consumption expenditure in households fell by 1.0% and a total of 10% over the full year.[3]

During the crisis it has also become clear that highly indebted households make economic recovery that much more difficult. Traditional fiscal and monetary policies such as increased government spending and lowering interest rates do not stimulate the economy as much since households need to use the increase in income to pay off debts. This weakens the stimulatory effects of such expansionary economic policies.

Fiscal policy: Governments can use fiscal polices – e.g. changing various tax rates or altering government spending – to influence aggregate demand. For example, an increase in **income taxes** (T_y) lowers the households' disposable income which in turn lowers consumption and aggregate demand. Increased **expenditure taxes**, e.g. VAT, have the same effect due to the decrease in real income.

- $\Delta\uparrow T_y \rightarrow \Delta\downarrow C \rightarrow \Delta\downarrow AD$ …or… $\Delta\downarrow T_y \rightarrow \Delta\uparrow C \rightarrow \Delta\uparrow AD$

---

1 Latest; October 2011

2 Possibly the most famous predictor of the impending crisis was Dr Nouriel Roubini (aka 'Dr Doom') of New York University who warned about the impending US housing bubble and economic effects as early as September 2006. See www.economicpredictions.org/who-predicted-the-financial-crisis.htm

3 See the US BEA at www.bea.gov, Table 2.3.1

Monetary policy: This is a macro policy tool in the hands of the Central Bank (or national bank) and involves adjusting **interest rates** (r) or changing the **supply of money**. 'Tight' monetary policy, here an increase in interest rates (r), has a twofold effect;

1. The opportunity cost of consumption rises since there is more interest foregone by keeping money in the bank; and

2. The cost of servicing loans (the interest payments) increases.

The compounded effect of an increase in interest rates will be that saving (S) increases and borrowing decreases. Both lead to a decrease in investment (I) and consumption(C) – these are components of aggregate demand in the economy. Conversely, 'loose' monetary policy of lowering interest rates will have a stimulatory effect on aggregate demand.

- $\Delta\uparrow r \rightarrow \Delta\uparrow S$ & $\Delta\downarrow I$ & $\Delta\downarrow C \rightarrow \Delta\downarrow AD$ ...or... $\Delta\downarrow r \rightarrow \Delta\downarrow S$ & $\Delta\uparrow I$ & $\Delta\uparrow C \rightarrow \Delta\uparrow AD$

Investment and AD

Profit and revenue expectations: Firms rely a great deal on expected expenditure in the economy for their planned investment. When firms expect future profits to rise they will be more willing to invest in the present. Accordingly, new technology which might serve to increase productivity and quickly regain investment costs will add to firms' willingness to invest. This was one of the reasons for the massive increase in Information Technology spending in the US during the latter part of the 1990s.

- $\Delta\uparrow$ profit expectations $\rightarrow \Delta\uparrow I \rightarrow \Delta\uparrow AD$ ... or ... $\Delta\downarrow$ profit expectations $\rightarrow \Delta\downarrow I \rightarrow \Delta\downarrow AD$

Policy expectations: All the main actors on the macro scene will adjust their behaviour when possible changes in government policies appear on the horizon. For example, the mere rumour of proposed interest rates hikes can cause a change in expenditure patterns when firms and households advance investment and consumption in order to avoid higher loan burdens in the future. An expectation of higher profit and income taxes will induce firms and households to cut back on expenditure in anticipation of falling net profits and incomes.

Government spending and AD

Since the government has the power to adjust government spending – which is a component of aggregate demand – this will have a direct impact on total expenditure in the economy. A boost in government spending will lead to an increase in aggregate demand and vice versa.

- $\Delta\uparrow G \rightarrow \Delta\uparrow AD$ ...or... $\Delta\downarrow G \rightarrow \Delta\downarrow AD$

Net exports and AD

As Section 4 will show, no economy is an island. Well, not literally, but in terms of being unconnected to other countries via trade and investment. There are three basic factors affecting an economy's aggregate demand due to international variables; exchange rates, trading partners' income and relative prices of trade partners. A fourth factor is of course barriers to trade such as import taxes (tariffs) and other restrictions, but assume for the time being that no trade barriers exist. I will exemplify with two trade countries, Argentina and USA. Assume that there is only one good, beef, and that it makes up a major proportion of each country's GDP.

Exchange rates: If the value of the domestic currency falls then imports will be more expensive for domestic households/firms – just as export goods will become cheaper for foreign households/firms. Assume that the Argentine peso (ARS) is at an exchange rate of **USD 0.23 for one ARS** (or USD 1 = **ARS 4.33**) and that a kg of beef costs $US 10 in America and ARS 43.3 in Argentina. The price – in terms of the exchange rate – is the same in both countries; ARS 43.34/4.33= $US 10.

Now posit that the Argentinean peso depreciates (falls) to ARS 1 = $US 0.20 (e.g. that the USD appreciates to $US 1 = ARS 5.00):

- The Americans experience that Argentinean beef has fallen in price and now need only $US 8.67 (ARS 43.34 x 0.20) to buy the 43.3 pesos necessary to buy a kg of Argentinean beef that still costs ARS43.3 since we are assuming *ceteris paribus*.

- The Argentines, conversely, will see how the price of American beef has risen to ARS 50.0 per kg. *Ceteris paribus*, one can expect Argentina's exports of beef to increase and the imports of American beef to decrease. Both variables are forces in increasing aggregate demand. (Note that this is a *shift* in aggregate demand not a movement along! It is the price of the **foreign**

currency that has changed – not the domestic price levels.)

- $\Delta\downarrow$exchange rate $\rightarrow \Delta\uparrow$X & $\Delta\downarrow$M$\rightarrow$ AD$\uparrow$ ... or ...
  $\Delta\uparrow$exchange rate $\rightarrow \Delta\downarrow$X & $\Delta\uparrow$M$\rightarrow \Delta\downarrow$AD

Trading partners' income: Posit instead that the exchange rate remains the same (along with everything else) but instead American incomes rise faster than Argentinean income. This will lead to increased consumption in America and as some of the goods consumed come from abroad, some of the increased consumption expenditure will go to Argentinean goods, which increases aggregate demand in Argentina.

- $\Delta\uparrow$trade partners' Y$\rightarrow \Delta\uparrow$X$\rightarrow \Delta\uparrow$AD...or...
  $\Delta\downarrow$trade partners' Y$\rightarrow \Delta\downarrow$X $\rightarrow \Delta\downarrow$AD

Relative prices: Finally, still assuming *ceteris paribus*, what if the price level in America increases more rapidly than in Argentina, i.e. that *relative prices* swerve in favour of Argentinian goods? Exactly; there would be an increased demand for Argentinean beef and therefore increased exports to the US. And, since US goods have become dearer across the board for Argentineans, imports from the US will fall. Both cause aggregate demand to increase in Argentina.

- $\Delta\uparrow$trade partners' price level $\rightarrow \Delta\uparrow$X & $\Delta\downarrow$M $\rightarrow \Delta\uparrow$AD ... or ... $\Delta\downarrow$trade partners' price level $\rightarrow \rightarrow \Delta\downarrow$X & $\Delta\uparrow$M$\rightarrow \Delta\downarrow$AD

Summary & revision

1. Aggregate demand is comprised of the **components** consumption expenditure, investment expenditure, government expenditure, export revenue and minus import expenditure.

2. Any **change in the economy** affecting one of these components will shift the AD curve.

3. **Consumption** is affected primarily by changes in interest rates, consumer confidence, expectations of future inflation, wealth (and perceived wealth), income, personal income tax rates and level of household indebtedness.

4. **Investment** (= changes in physical capital in firms) will be affected by interest rates, business confidence, profit taxes (also called corporate taxes) and level of corporate debt.

5. **Government spending** is a political decision whereby the government's decision to change spending on infrastructure and public services affects aggregate demand.

6. **Export revenue and import expenditure** is affected by exchange rates, trade partners' income, trade barriers, relative prices and relative inflation.

MACROECONOMICS

42. Aggregate supply

Key concepts:

- Aggregates supply (AS) and output
- The upward sloping AS curve – short run aggregate supply (SRAS)
- Shifts in SRAS

Aggregate supply (AS) and output

The aggregate supply curve shows the ability and propensity of firms to put goods on the market, or the relationship between the price level and real output. In the short run, the higher the price level, the more goods firms will be able and willing to supply – the short run aggregate supply curve will thus show positive correlation between the price level and real national income. In the long run, it is not as simple, since there is a great deal of controversy in the theoretical divide between two broad schools of thought.

> **Definition: 'Aggregate supply'**
>
> **Aggregate supply** is the planned output of goods and services in an economy during a period of time at different price levels. There is positive correlation between the price level and aggregate supply. (Note that we assume that factor prices remain unchanged along the AS curve.)

The upward sloping AS curve – short run aggregate supply (SRAS)

Production takes place because suppliers anticipate revenue and profit from coming sales. Costs, on the other hand, rise simultaneously with suppliers' increase in output. Suppliers run a risk of not receiving revenue enough to match costs, since they cannot predict future prices. Therefore, should real profits exceed expected profits suppliers will have an incentive to increase output, which strongly suggests an upward sloping aggregate supply curve. Assuming that the prices of all factors of production (e.g. wage rates, raw material prices, price of capital, etc) remain constant, then suppliers will indeed increase output when the price level rises.

This is portrayed in Figure 42.1 as the upward-sloping portion of the *short run aggregate supply* (SRAS) curve. The aggregate

supply of an economy in the short run is based on the notion of fixed factor prices and availability – e.g. the constraints are similar to those used in Chapter 4, that is to say, that all else remains constant – the *ceteris paribus* assumption.
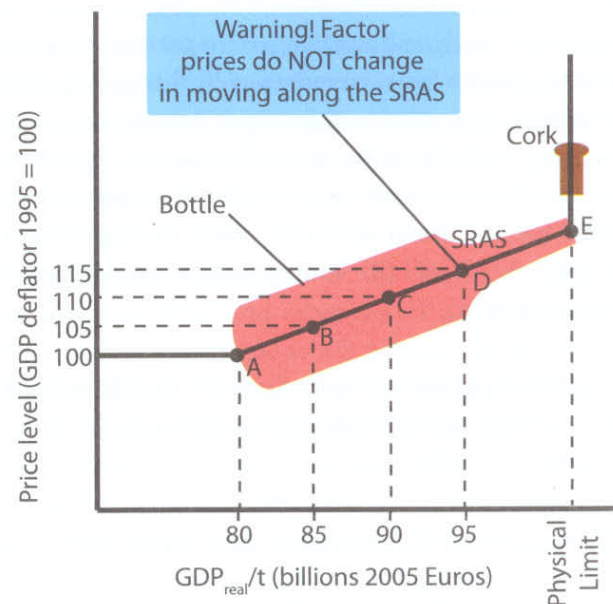
> Higher market prices will induce suppliers to increase output. The AS curve is drawn under the assumption that prices of factor inputs do not change in the short run, so a higher price level means larger profit margins for firms. Thus the price level and output are positively correlated. The upward slope of AS also shows how firms experience diminishing returns (as existing fixed factors are ever increasingly utilised) and increasing scarcity of variable factors. Both serve to create bottlenecks in production. (Shown by the shadowed bottle in the figure.)
> When it becomes impossible to increase output in the short run, the physical limit of the economy's capacity is reached and real GDP is no longer correlated to the price level. An increase in the price level simply means a general rise in prices without any increase in output. (Shown by the 'corked' portion of the SRAS curve.)

Figure 42.1 Short run aggregate supply (SRAS)

MACROECONOMICS

HL; What holds true for a single firm will in this particular case hold true for the aggregate, namely that increased output will result in higher marginal costs – the sum of MC curves for individual firms will give us the SRAS curve for the economy as a whole. This helps explain the relatively high elasticity of aggregate supply in the short run.

Let's look at the three sections of the SRAS curve in Figure 42.1 in turn:

- **Horizontal section**: The horizontal portion of the short run aggregate supply curve shows output when the economy is in severe recession, or *depression*. Output can increase without an increase in price, i.e. short run aggregate supply is horizontal and therefore uncorrelated to the price level. Firms can increase production regardless of the price level since there is:

  o High unemployment and thus abundant labour and inability to bid up wages – no cost pressure on firms due to increasing wages

  o Excess capacity in firms, which enables them to increase output without incurring higher (marginal) costs

  o Low demand in the economy which makes firms unwilling/unable to raise their prices

- **Upward sloping section**: At an output beyond €80 billion (**point A** and beyond on the short run aggregate supply curve) firms will increase output in accordance with a rise in final prices (= market prices) but will start to encounter *diminishing returns* and *supply constraints*.

  o Fixed factors of production such as machines, assembly lines and factory space will not increase during the short run; *diminishing returns* set in.

  o Adding to this is the increased *scarcity of factors* as output increases in the short run, leading to *bottlenecks* – supply constraints – in production; increased scarcity of skilled labour, transportation, delivery and availability of raw materials will all serve to limit firms' ability to put more goods on the market. (Remember, factor prices do not change, so increasing production costs are the result of overtime, increasing maintenance costs of machines, search costs for ever scarcer raw materials etc.)

  o As firms' costs increase there must be an incentive of higher (final) prices for firms to be willing and able to increase output. **Points A through E** show the positive correlation between the price level and real GDP.

(The bottle in the diagram is my incredibly childish way of showing that the upward slope is due to bottlenecks in supply and **NOT** higher factor prices.)

- **Vertical section**: The vertical portion of the short run aggregate supply curve (beyond E) illustrates how the economy ultimately hits the output capacity ceiling. This is the *physical limit* of firms' short run output capacity. Up to **point E** the economy is able to increase output by paying workers overtime, hiring ever more labour and pushing capital to the limit. At some point (here E in the diagram) workers will be reluctant to work overtime, additional labour will be unavailable, and firms will be unwilling to drive machines into the ground without major overhauls (= repairs and maintenance). The inducement of higher final prices will not increase output in firms no matter what the profit margins are, so the short run aggregate supply curve becomes vertical – or completely inelastic.

(The cork is a silly illustration showing that no more 'flow' – e.g. output – is possible at any point beyond E'.)

A final technical note: I have drawn the shape of the short run aggregate supply curve (SRAS) in Figure 42.1 partially to show the historical link to Keynesian theory (see Chapter 43) and partially to illustrate the issue of how an economy will show increasing difficulties in increasing output due to supply bottlenecks. In the continuation, I will confine the range of the short run aggregate supply curve to the upward-sloping section, e.g. **points A to E**. It is within the span of the upward-sloping section that all short run analysis will take place and the horizontal and vertical sections add little in terms of analytical value. This is also in keeping with what has become standard 'mainstream' usage of short run aggregate supply curves.

MACROECONOMICS

287

Shifts in SRAS

The short run aggregate supply curve in the AS-AD model will shift when production costs for firms change. Three specific short run influences can be identified:

1. **Price of labour**: An increase in wage rates will mean higher production costs for firms and shift the short run aggregate supply curve to the left, from $SRAS_0$ to $SRAS_2$ in Figure 42.2. This will result in lower output, Y_2, and a higher price level, P_2. A decrease in the wage rate will naturally have the opposite effect.

2. **Price of inputs**: Changes in other factor markets, e.g. the markets for raw material, capital and components will have effects on the cost picture of industrial firms. For example, a decrease in the price of steel will enable producers of cars, houses and washing machines etc to increase output at all price levels. This is shown by a shift in short run aggregate supply from $SRAS_0$ to $SRAS_1$, increasing output to Y_1 and lowering the price level to P_1.

3. **Taxation and legislation**: There are a variety of taxes levied on production which increase costs for firms. Labour taxes (a percentage on wages paid by employers which go to social security contributions, pensions and such) add to overall labour costs. Profit taxes (or corporate taxes) deny firms money that could be used for investment. Environmental taxes on emissions add to firms' costs. Legislation on minimum wage rates; overtime regulation; work hours etc also add to the total cost picture of firms. Increased taxes on labour etc and stricter regulatory legislation will shift short run aggregate supply to the left, from $SRAS_0$ to $SRAS_2$, while lower taxes and looser legislation will shift the short run aggregate supply curve to the right, from $SRAS_0$ to $SRAS_1$.

In addition to factors affecting the *price* and *availability* of factors of production, there are factors such as the *efficiency* of factors and *external shocks*. Improvements in technology, labour and production methods will increase aggregate supply while severe storms and/or natural disasters can decrease aggregate supply.
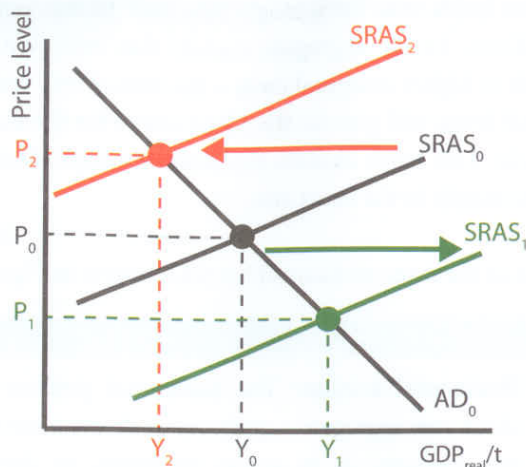

Figure 42.2 Increase and decrease in short run aggregate supply

Summary & revision

1. **Aggregate supply** is the planned output of goods and services in an economy during a period of time. The short run aggregate supply curve is shows positive correlation between the price level and planned output.

2. A **key assumption** of the SRAS curve is that factor prices and factor availability are considered constant. Any change in these variables will shift the SRAS curve.

3. **Shifts in short run aggregate supply** are primarily caused by the price, availability, efficiency and quantity of factors of production. Short run aggregate supply shifts due to:

   a. changes in **factor prices** (wages, raw material, capital)

   b. changes in factor efficiency (new technology, production methods)

   c. taxation (profit taxes, labour taxes, environmental taxes)

   d. government regulations (overtime regulation, subsidies)

   e. external shocks (natural disasters, weather)

MACROECONOMICS

43. Keynesian vs. Monetarist/New Classical View of LRAS

Key concepts:

- Monetarist/new classical model of LRAS
- Keynesian model of AS

I live in an 'Expatriate Bubble' in south Jakarta Indonesia at about 50 meters above sea level – with the shore line some 40 kms away. If one were to take all the books written about John Maynard Keynes and dump them into the sea, I'd have beach-front property and Al Gore would be howling "I told you so!" However, books on Keynes are of high marginal social benefit, so book-dumping would be a major loss.[1] In any case, *macroeconomics* – as a specific field of economics – was basically invented by Keynes (see *Celebrity Bubble* following) during the 1920s and '30s. In order to understand the 'competing' schools of thought on the subject of demand-side or supply-side, it is necessary to briefly look at the historical background of macroeconomics and trace the strong theoretical and ideological differences which come to a point in the dispute over the shape of the long run aggregate supply curve.

Keynesian model of AS

Conventional economic wisdom in the pre-Keynesian era held that both goods and labour markets ultimately clear and that market forces would ultimately do away with any excess supply or demand. Economic analysis in the late 19th and early 20th century, the **new classical school**, was primarily *microeconomic*, where the focus was on individual markets and the market clearing functions of supply and demand. This line of thinking was equally applied to labour markets; labour was no different from any other commodity and as long as the labour market was kept free from interventionist forces (more on this in Section 3.4) there would be full employment. One could say that the new classical school sees no discernible difference between the market for chickens and the market for chicken-pluckers. In other words, the labour market would always tend towards equilibrium in the long run.

The Great Depression of the 1930s and prolonged mass unemployment was viewed by new classical theory as the result of market interventionist and anti-competitive forces such as trade-unions which kept wages above market clearing levels. As long as wage rates responded to market forces and were flexible the labour force would adapt to wage changes. According to the predominant view, unemployment was 'voluntary' to the extent that labourers were not willing to accept jobs at the going wage rate.

This view came under scrutiny and subsequent challenge during the 1930s, especially when unemployment rates hit hitherto unseen levels – over 15% in Great Britain and 25% in the US – in the early 1930s and remained high. Two simultaneous – and largely unrelated – schools of thought developed; one in Sweden and one in England. The Swedish School, later known as the 'Stockholm School' evolved primarily from the work of Gunnar Myrdal and Bertil Ohlin (later a Nobel laureate) while in England it was Keynes who was the prime instigator.[2] The basic position taken in this new line of theory was that the drawn-out depression where high unemployment and stagnant (or falling) output levels existed for such long periods could neither be explained nor solved by the predominant theory of the day.

---

1 Don't worry, I'm waiting for global warming and melting ice caps to take care of things nicely. Here at the Shady Oaks Retirement Home I will be fishing for marlin from my balcony.

2 *Makroekonomi – teori, politik och institutioner*, Klas Fregert & Lars Jonung, Studentlitteratur 2003, pages 105 – 120 (The authors point out that the Swedes' very similar theory predated Keynes by a few years and posit that had Myrdal, Ohlin and others not written in Swedish we could very well have been talking about Stockholmian theory rather than Keynesian theory today.)

MACROECONOMICS

CELEBRITY BUBBLE: JOHN MAYNARD KEYNES (1883–1946)

John Maynard Keynes (rhymes with 'rains' and Keynesian rhymes with 'rains-ian' – whatever else you might hear) was in all probability one of the most brilliant and gifted people ever to walk the face of the earth. Luckily, he also became one of the most influential. What else can you say about someone who basically 'slipped' into economics and thereby created a new science? Who, according to legend, made his fortune by reading the morning paper in bed every morning and then placing a call to his stockbroker? Who had a heavy hand in creating several of the most influential international agreements in the post-WWII period? Who ... I give up. It is impossible to fit the gigantic figure of Mr Keynes into a few hundred words. Here are a few basic biographical notes topped off by five books well worth reading.

Victorian in upbringing and Old School to the hilt, Keynes displayed brilliance from the first. At 14 he won a scholarship to Eton and went on to King's College at Cambridge. There he studied mathematics but was begged by the famous Alfred Marshall to become an economist. Instead, he went to India for two years as a civil servant, after passing the civil service examination with the second highest marks – his lowest marks being in economics: 'I evidently knew more about Economy than my examiners.'

Having accomplished little in India other than getting a pedigreed bull sent to Bombay, as he later remarked, he returned to Cambridge where he taught and also ran the prestigious Economic Journal. He became a part of the group of artists and intellectuals know as the Bloomsbury Group, including Leonard and Virginia Woolf, E. M. Forster, and Lytton Strachey with whom he had one of several homosexual affairs.

He speculated in shares and bonds, making the considerable fortune of some £500,000 in current values. He ran the Cambridge theatre and managed the restaurant – plotting food receipts against what was currently playing to see which plays could be linked to increased demand for food. He married a famous Russian ballerina, Lydia Lopokov. In between all this he reshaped the world.

He joined the Treasury at the outbreak of WWI and acted as an advisor during the Versailles negotiations at the end of the war – only to return in disgust at the terms imposed on an impoverished Germany and subsequently to write his first major piece 'The Economic Consequences of the Peace on the economic consequences of the Versailles Peace Treaty', in which he basically predicted WWII. In 1934 he visited President Roosevelt in Washington and saw the ongoing New Deal of public sector job creation – and urged Roosevelt to extend the program.

In a letter to George Bernard Shaw in 1935, Keynes wrote that 'I believe myself to be writing a book on economic theory which will largely revolutionize ... the way the world thinks about economic problems'. * He was right – the following year he published The General Theory of Employment, Interest and Money. This incredibly dry and theoretically intimidating work would turn economic theory on end, create the field of macroeconomics, and shape economicpolicy for the entire post-WWII period.

Keynesian critique of new classical thinking

"The real difficulty in changing any enterprise lies not in developing new ideas, but in escaping from the old ones". The origins of Keynesian theory are squarely rooted in the seeming failures of depression era economic thinking which did not seem to be able to deal with stagnant growth and high unemployment. Here are the main points of Keynes's criticism of the new classical view:

- Markets are inherently (= essentially) unstable and do not necessarily clear. It is quite possible for labour markets to render high levels of unemployment even in the long run. Keynes harshly criticised new classical thinking for **fallacy of composition**; simply because individual markets might clear does not mean that this holds true for the aggregate. Lowering wages might well increase the demand for – and amount of – labourers in one industry but not in *all* industries.

- Wages are '**downward sticky**', i.e. labourers/unions are most unwilling to accept cuts in wage rates. This adds to labour market disequilibrium by keeping wages above market clearing level. Real wages simply do not

fall enough to completely clear the market and restore full employment. (Revise 'minimum price' in Section 2.1.)

- Since **labour markets are imperfect**, non-interventionist policies do not help unemployment and in fact may serve to keep unemployment rates high over long periods of time.

The strength in Keynes's work lay in his predisposition for hands-on applicability. While his work was heavily based in mathematics – he was, in fact, a mathematician by education – he was clearly focused on practical solutions. The logical conclusion to the criticism of prevailing new classical thought was that there is merit in **government intervention** to help create market equilibrium. Fiscal policies such as increased government spending can help achieve full employment by increasing aggregate demand. According to traditional Keynesian theory, disequilibrium on the labour market is primarily due to lack of demand for labour – this is *demand deficient unemployment* – and can be remedied by increasing aggregate demand and therefore increasing the demand for labour. (Remember 'derived demand'?!) While 'pre-Keynesian' thought accepted Say's Law of '...supply creating its own demand...'[3], meaning that there could never be a permanent glut in supply (and thus deficiency in demand), Keynes vehemently disagreed and one might well imagine him reversing the order; '...demand creates its own supply...'.

The Keynesian aggregate supply curve – 'In the long run we are dead'

I bet the caption caught your eye. You will understand it in a moment. Figure 42.1 shows the Keynesian aggregate supply curves which ranges from a horizontal portion, a 'trade-off' portion and a vertical portion. (Note that the Y-axis reads 'Price level' rather than 'GDP deflator', which will be the case in all diagrams when referring to 'P_0, P_1' rather than actual index values.)

The 'traditional Keynesian' aggregate supply curve, AS, shows three possible ranges of output:

- **Horizontal portion**: The depression range of mass unemployment (up to Y_0) follows the course of the previous aggregate supply curve, where output increases without an increase in the price level. The horizontal portion of the 'inverted-L' shaped curve

shows that income can increase to Y_0 without a rise in the price level. This is based on:

- o Keynes's premise of *downward stickiness of labour*. Since labourers/unions are unwilling to accept lower wages (this is the 'downward stickiness of labour prices' part[4]) during a period of demand deficiency and resulting unemployment, a situation of high unemployment may persist in the long run.

- o Due to the *abundance* – excess supply – *of labour* and other factors at low income levels, labourers will be highly reluctant to bid up wages even as output increases. Imagine that output increases from any output level up to Y_1 with the ensuing increase in demand for labour. Both newly hired and existing workers would not be in a position to bid up wages since there are thousands waiting in line for their jobs. Thus it is possible to increase output without creating upward pressure on wages and final prices, resulting in a horizontal aggregate supply curve.

- **Trade-off portion**: At Y_0 the economy is moving closer to the full employment level of output and firms are responding to higher final prices for their goods and services. This middle range portrays the correlation between a higher price level and increasing real output (explained earlier) leading the aggregate supply curve to be upward-sloping. This range, Y_0 to Y_{FE}, illustrates two important points.

- o The first is that aggregate supply will increase only when firms are able to enjoy higher final prices for their output and thus cover additional costs arising from bottlenecks in supply.

- o This portion highlights one of the most important policy debates in economics; the apparent *trade-off between unemployment and inflation*. Recall that one of the key conclusions of Keynesian theory is that markets are imperfect and thus government intervention

---

3 Jean-Baptiste Say, famous French *laissez faire* advocate and new classical economist, (1776-1832)

4 I often explain 'downward sticky' by referring to good and bad wines. It's easy to get used to good wines – and most difficult to move back down on the quality list once one is used to the good stuff. Acquired taste is thus 'downward sticky'.

MACROECONOMICS

is necessary in order to create labour market clearing and thus full employment. The middle-range of the aggregate supply curve indicates that governments will face a macroeconomic opportunity cost issue; increased government spending (which stimulates AD) might result in lower unemployment at the cost of a higher price level, i.e. inflation. This trade-off between inflation and unemployment is a central subject of Chapter 54, the Phillips curve.

- **Vertical portion**: The vertical range of AS at Y_{FE} and beyond is the same as the physical limit illustrated earlier (Figure 43.1 SRAS) where firms simply cannot increase output whatever the incentives of final prices. The complete price inelasticity of supply beyond point Y_{FE} illustrates the effect of ever-more scarce factors of production and thus increasing output constraints. This is the *full employment level of output*, Y_{FE}. Firms will not be able to hire additional labour and no increase in output is possible; the aggregate supply curve is vertical. Any increase in aggregate demand will be purely inflationary.
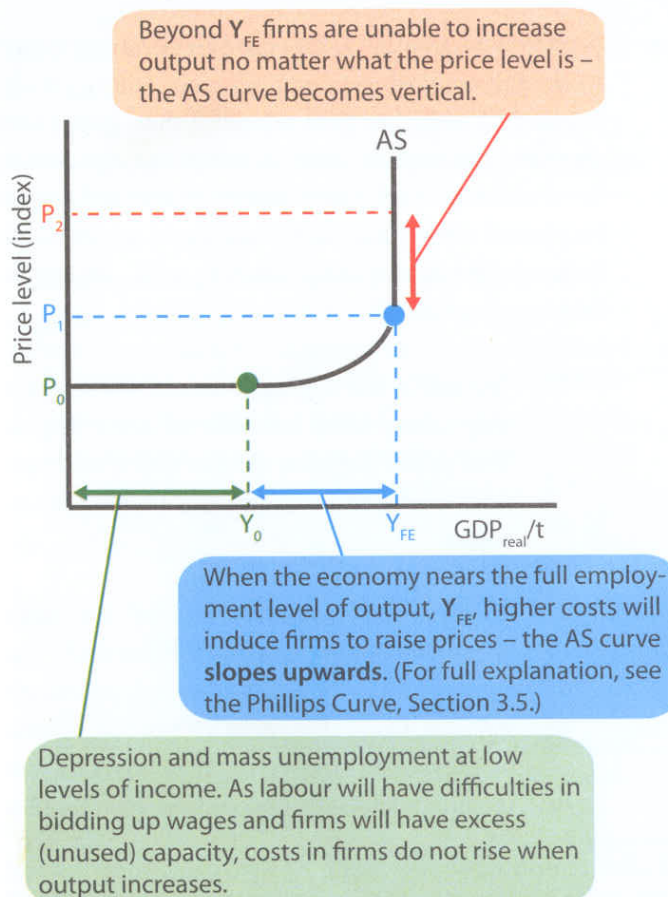
Beyond Y_{FE} firms are unable to increase output no matter what the price level is – the AS curve becomes vertical.

When the economy nears the full employment level of output, Y_{FE}, higher costs will induce firms to raise prices – the AS curve **slopes upwards**. (For full explanation, see the Phillips Curve, Section 3.5.)

Depression and mass unemployment at low levels of income. As labour will have difficulties in bidding up wages and firms will have excess (unused) capacity, costs in firms do not rise when output increases.

Figure 43.1: Keynesian aggregate supply

SR and LR

We finish where we began, where we all die. Have you noticed an inconsistency between the current syllabus heading and the curves in Figure 43.1 - *Keynesian aggregate supply* above? Have a quick look and ponder before you read on. Did you see it? The heading reads '*long run aggregate supply*' but the curve in the diagram is labelled 'aggregate supply'. This is completely in line with Keynesian reasoning on aggregate output, where it may take years, or even decades for labour markets to clear, during which the economy can remain at an equilibrium below full employment – since there is nothing inherent in the economic system to move the economy out of depression.

The Keynesian prescription of increased government spending during recession/depression to stimulate the economy ('priming the pump' as it was known during Keynes's time) serves to increase output and rectify disequilibrium on the labour market. In taking care of the short run, the long run will take care of itself, so to speak. So, while it might be possible in the *long run* for real wages to fall enough to create full employment without government intervention …in the long run we are all dead.

For the record; Aggregate supply according to 'Keynesians' and Keynes

There is no end of confusion as to what Keynes actually did or did not say. Much of what is considered to be 'traditional Keynesianism' was in fact the result of followers of Keynes rather than the master's own work. Part of the confusion apparently arises from the fact that Keynes was by training a mathematician rather than an economist, leading him to use mathematical formulae for his illustrations, rather than diagrams (against the express recommendations of his famous teacher, Alfred Marshall, to whom Keynes lovingly referred as "an absurd old man"). Most of the diagrammatical work was done by others, primarily the economists (and Nobel laureates) Paul Samuelson and Sir John Hicks. In any case, while Keynesian theory has commonly utilised **diagram B** following, it is in fact highly likely that Keynes himself accepted the standard new classical theory of an upward-sloping aggregate supply curve (**diagram C**) given his assumption of constant nominal wage rates making it possible for firms to increase output when final prices of goods increase since real wage rates will have effectively fallen.
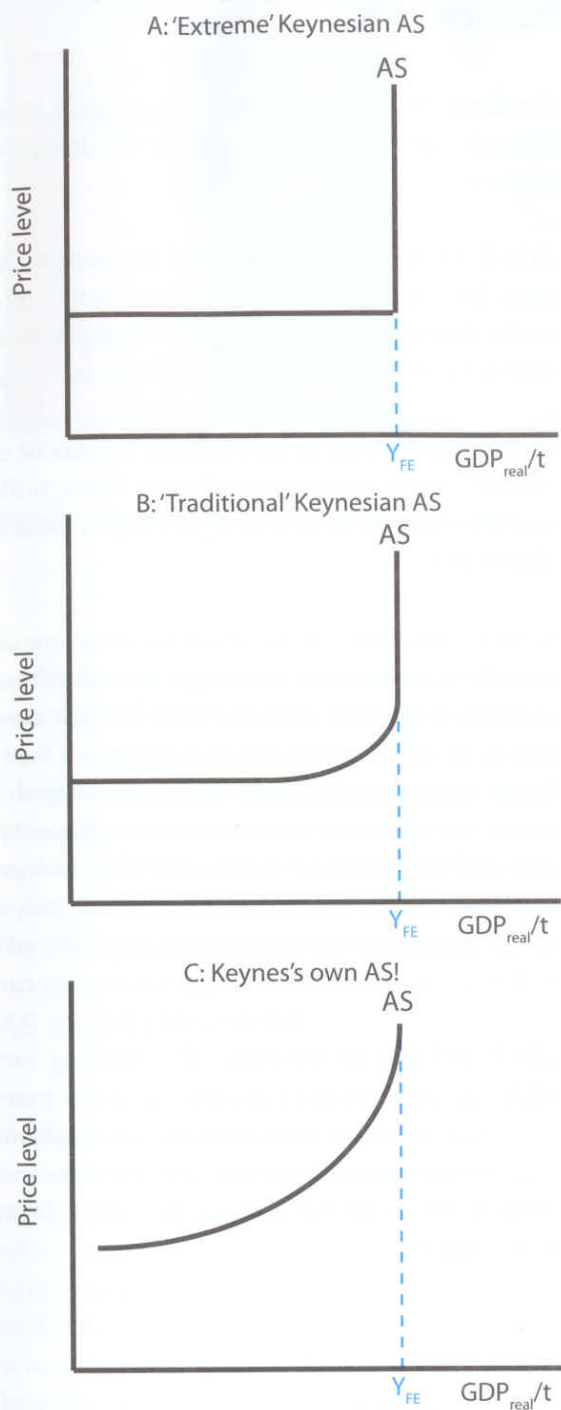
A: 'Extreme' Keynesian AS

B: 'Traditional' Keynesian AS

C: Keynes's own AS!

Figure 43.2

Monetarist/new classical model of LRAS

The Keynesian view of aggregate supply dominated economic thinking and policies throughout the 1950s and '60s, whereupon a counterview arose which took as its base the new classical premise of perfectly functioning markets. This was the **monetarist school** of thought initially put forward by Milton Friedman (see *Celebrity Bubble* following) which would also spawn a number of similar schools of thought, primarily the new-classical and supply-side schools (see 'A little depth...' below).[5] The monetarist line of reasoning took issue with a number of key Keynesian assumptions:

Monetarist criticism of Keynesian AS

1. The starting point for the monetarist school was that 'People are not fooled by having more money'. In other words, as income increases when moving along the short run aggregate supply curve people would realise that the higher price level would hollow out their – unchanged – wages. Labourers in the monetarist/new classical view thus do not suffer from '*money illusion*' but are well aware of the negative effects of a higher price level on *real wages*.

2. This view ('new-classical' from now on) therefore strongly disagreed with the Keynesian assumption that wage rates would remain unchanged in the long run. Wages are market-based and therefore highly flexible - workers' **inflationary expectations** (see Section 3.5) would force them to use their bargaining power to bid up wages when the price level increased in order to retain their purchasing power.

3. Since higher wages 'eat up the distance' between the final price firms get for their goods and the costs of producing them, the short run aggregate supply curve will shift to the left when wage levels in the economy are bid up. (This is no different from saying that an increase in factor prices decreases supply, i.e. *aggregate* supply.) Thus there will be a **separate short run aggregate supply curve** for every wage and factor price rate.

5 There are a number of distinctions between these schools, but I will limit myself to using 'monetarist' or 'new classical', since the nuances are too far beyond the scope of the syllabus.

CELEBRITY BUBBLE: MILTON FRIEDMAN (1912 - 2006)

Milton Friedman was perhaps one of the foremost representatives amongst economists advocating limited government intervention. His close to 40 years of research and teaching at the University of Chicago made him the pivotal figure in what is now known as the 'Chicago School', which is based on free markets and stable money supply being the optimal method of attaining macroeconomic growth and stability. He is also the most famous economist of recent times, and the central figure in the monetarist school.

His work mostly focused on price theory, where the core theme is how individual market prices are set. His theoretical direction took its vantage point in the neo-neo-classical centrepiece quantity theory of money – which proposes that price levels and inflation are primarily dependent on the money supply. His books *Studies in the Quantity Theory of Money* from 1956 and *Monetary History of the United States* (coauthored with Anna Schwartz) argued that increased monetary growth would not in the long run increase real output but only the price level – which set the stage for monetary theory during the 1970s as much of what Friedman had predicted seemed to come true.

The early and mid 1970s was a period of high inflation and low – or falling – growth, so called 'stagflation' or 'slumpflation'. This was the period where the Phillips-curve relationship seemingly broke down (Section 3.5) and Friedman put forward the expectations augmented Phillips curve – which ultimately resulted in the New Neo-classical concept of a natural rate of unemployment and the long run supply curve.

Friedman was an informal economic advisor to the American presidents Nixon and Reagan and won enough awards to fill two more pages – but one can assume that his winning of the Nobel Prize in economics in 1976 is worth mentioning. While he has always been controversial, there is no denying his enormous impact on economic theory. Indeed, economic policy too, as the entire Reagan and Thatcher 'epochs' of the 1980s were heavily influenced by Friedman's work.

New-classical LRAS

Putting the above into points in Figure 43.3, we get a long run aggregate supply curve showing that real GDP is independent of the price level.

- **A to B**: In the new classical model any increase in the price level accompanying an increase output, **A to B** in the diagram, will result in higher wages for workers who act to at least retain (if not increase) real wages.

- **B to C**: The increase in factor prices, e.g. labour costs, shifts short run aggregate supply from $SRAS_0$ to $SRAS_1$ and the economy returns to Y_0 at a higher price level, shown by C.

- **C to E**: The long run aggregate supply curve shows how firms' costs adapt to changes in final prices. For example, if the price level increases by 5% and wages and other factor prices also increase by 5% then real factor costs and real wages have not changed. This means that there is in fact no increase in the real price level and therefore no real inducement for an increase in output. Further movements along $SRAS_1$ will result in the same thing, moving the economy from point C to D to E via a new short run aggregate supply curve.

- **LRAS and full employment**: The resulting vertical LRAS curve shows the economy's long run potential GDP. At this level of output there is full employment; Y_{FE}. (**Strong warning here**: this does NOT mean '*zero* unemployment' but the '*natural* rate of unemployment' as we shall see.)
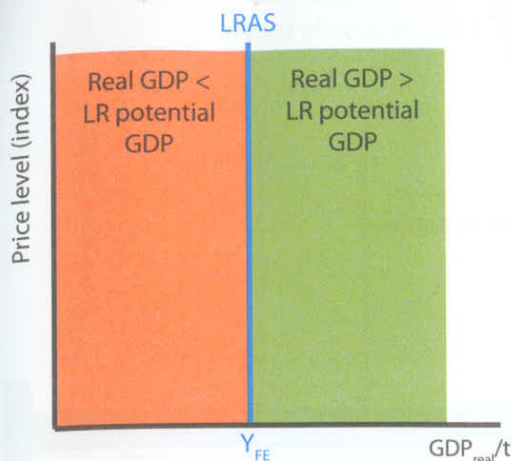
Figure: 43.3 Monetarist/new-classical LRAS

Looking at it in the reverse order, starting at **point E**, consider how a *fall* in output prices at a given money wage rate would increase the real wage rate and therefore increase the supply of labour – causing an excess supply of labour. Since wages are flexible they will quickly decrease and restore equilibrium on the labour market – and the lower real wage rate would increase aggregate supply from **SRAS₂** to **SRAS₁**, moving the economy from **point E** to **point C** in the diagram. The long run aggregate supply curve is thus uncorrelated to a change in the price level.

LRAS vs. physical limit

It is most important that you do not confuse long run aggregate supply with the physical limit outlined earlier. The LRAS curve shows *potential real output in the long run*, which is the same as the full employment level. It is possible for real output to exceed this point in the short run – according to most, but not all, classically orientated economists – just as it is possible for real GDP to be less than potential output. The key is in understanding that 'potential' in this context indicates the long run capacity of firms in the economy to remain at a certain output level before having to adjust to changes in factor markets. So while firms can physically produce beyond Y_{FE} in Figure 43.3 (right diagram) above, increasing factor prices will force them back to the output level on the long run aggregate supply curve.

Aggregate Supply

1. How/why can an economy be in equilibrium at less than full employment according to Keynes?

2. What was the main criticism of standard economic policy put forward in the 'Keynesian revolution'?

3. Why might there be unemployment at the equilibrium level of income according to Keynes?

4. The monetarist school in its turn put forward criticism of Keynesian assumptions. Explain how this led to a short run aggregate supply curve and a vertical long run aggregate supply curve.

5. "The physical output limit is not the same as long run aggregate supply". Explain.

Summary & revision

1. The **Keynesian aggregate supply curve** shows three sections;

   a. a **horizontal section** with high unemployment, excess capacity and unsold stocks in firms;

   b. a **'trade-off' section** where bottlenecks in supply together with diminishing returns render an upward-sloping curve;

   c. a **vertical section** showing the physical limit of the economy

2. The **monetarist/new classical view** is that there is a long run aggregate supply curve showing the long run *potential* output of the economy. This is the **full employment level of output**. The vertical LRAS curve in this model is based on the premise that any increase in output beyond income at full employment will drive up factor prices and decrease short run aggregate supply.

MACROECONOMICS

295

44. Shifting LRAS

Key concepts:

- Factors causing a shift in LRAS
- Effects of LRAS shifts according to the monetarist/new classical and Keynesian model

Factors causing a shift in LRAS

In Chapter 42 the issue was short run aggregate supply – SRAS – and three main variables affecting SRAS were identified; price of labour, price of other factor inputs such as raw materials, and taxation/legislation. While all of these indeed affect the long run, there are several rather specific **long run influences** for LRAS:

- New *oil/mineral finds*, improved *pesticides*, better *land* **use**, superior *high-yielding grain* varieties, improvements in *farming methods*

- Improvements in the *labour force* such as better education standards, *immigration* and *population* growth, *decreased income taxes* incentivising households to work more

- Introduction of *new technology* which increases productivity levels in an economy, new *production methods* (again, an increase in productivity), change in *tax regulations* that encourage firms to invest, *deregulation* and *privatisation* increase capital stock in an economy

The list can be rather lengthy. Basically anything in an economy that increases the **quantity and quality of factors of production** will shift the LRAS. In Chapters 60 to 62 we will return to the issue of how governments can intentionally change polices in order to influence long run aggregate supply – so called **supply-side policies**.

Effects of LRAS shifts according to the monetarist/new classical and Keynesian model

New classical model: Long run aggregate supply – again, representing long run *potential* output – will be affected primarily by changes in the *quality* and/or *quantity* of factors of production. However, long run aggregate supply will NOT be affected by a change in the *price* of factors since the AS-AD model is built on the assumption that factor prices will adjust to any increase in the price of final output. The long run aggregate supply curve is vertical for this very reason; it shows potential output regardless of price level. The tenet (= basic core belief) of the new classical school of *LRAS being uncorrelated to price* has serious implications for policy views put forward by supply-side economists as shall be seen in Chapter 60.
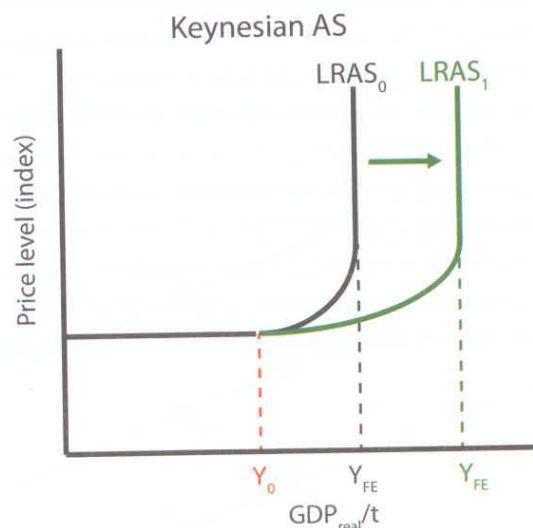

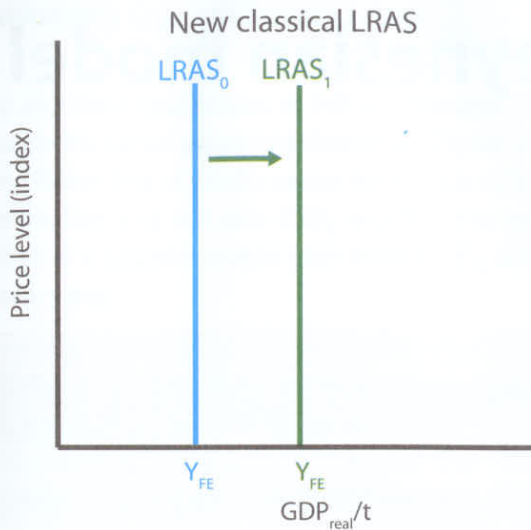Figure 44.1a – shifts in LRAS in the Keynesian and new classical model

New classical LRAS

Figure 44.1ab – shifts in LRAS in the Keynesian and new classical model

Summary & revision

1. All the **variables** affecting SRAS will affect LRAS – *except the price* of factors of production. We assume all factor prices to remain unchanged *along* the SRAS curve.

2. Anything which affects the **quality and quantity of factors** of production will shift LRAS.

3. Main influences on **LRAS**:

   a. **Land:** increases in arable land, improvements in farming techniques, oil and mineral finds

   b. **Labour:** deregulation of labour markets, increase in population or labour force, lower income taxes

   c. **Capital:** new technology, increased investment, deregulation

LRAS will shift due to a change in the quality and price of factors of production – BUT NOT DUE TO a change in the price of factors. Variables which are specific to long run aggregate supply involve potential productivity changes due to changes in the quantity and/or quality of factors, for example changes in labour supply, technology, production methods and labour capital.

Keynesian model: The change in LR factors affecting aggregate supply cause a similar shift in the LRAS curve. However, note that while the Keynesian version also causes the income at full employment level to increase, there is a 'portion' of the LRAS curve below Y_0 which in fact doesn't change at all. This has implications for Keynesian policy responses to severe recessions as shall be seen in Chapter 62.
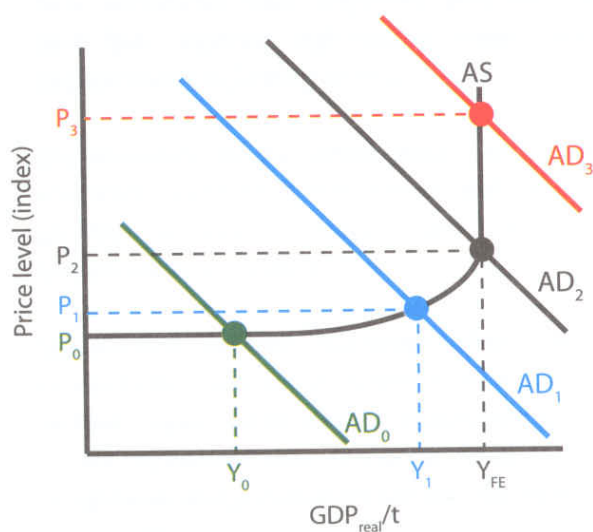
45. Equilibrium in the Keynesian model

Key concepts:

- Equilibrium; planned expenditure equals planned output
- Deflationary (recessionary) gap
- Inflationary (expansionary) gap
- Effects of an increase in AD at low and high levels of income
- Possibility of equilibrium below full employment in the long run

Equilibrium; planned expenditure (AD) equals planned output (AS)

Aggregate demand is comprised of total planned expenditure in an economy by households, government, firms and the foreign sector. The level of income where planned expenditure equals planned output is equilibrium output.

When planned expenditure (AD) equals planned output (AS), the economy is in equilibrium (AD= AS). Different levels of aggregate demand show different levels of equilibrium national income.

Figure 45.1 – Macro equilibrium in the Keynesian AS-AD model

In the Keynesian model, there is no distinction between the long run and short run so macroeconomic equilibrium is possible at all levels of income. Y_1 to Y_{FE} in Figure 45.1 show four different possibilities:

- Y_0 shows the economy in recession – low levels of national income, high unemployment and excess capacity in the economy. Clearly not politically desirable and most likely to lead to government intervention to stimulate aggregate demand.

- Assuming, for the sake of simplicity, that government has implemented stimulatory measures such as increased government spending, aggregate demand has increased to AD_1 and the economy has moved towards the full employment level of income – income is rising and unemployment is falling. As factors become increasingly scarce and supply-side bottle-necks set in, inflation rises also.

- At Y_{FE} all available factors are being used and any further increase in aggregate demand – to AD_3 and beyond – is purely inflationary. Output capacity simply cannot meet demand beyond Y_{FE} as no additional factors can be employed.

Deflationary (recessionary) gap

Assume an initial equilibrium at full employment (Y_{FE}) and that households expect property prices to fall (Figure 45.2). The perceived future loss of wealth causes households to hold back on consumption and AD falls (AD_0 to AD_1). The gap (often referred to as a negative output gap) between Y_{FE} and Y_1 is a **deflationary gap**.
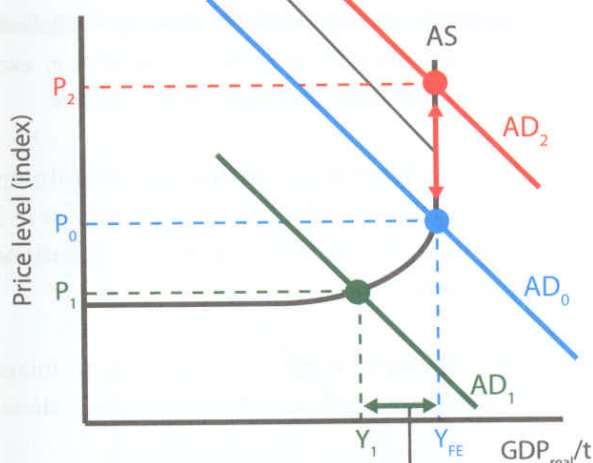
> **Definition: 'Deflationary' or 'recessionary' gap**
>
> When de facto output in terms of macro equilibrium is below potential output, there is a negative output gap – a **deflationary gap**.

> **Inflationary gap:** When AD rises at the full employment level of GDP, the result is inflation but no increase in output. The distance between P_0 and P_2 is an inflationary gap.

> **Deflationary gap:** At any point of macro equilibrium below the full employment level of output there will be a deflationary gap – here between Y_1 and Y_{FE}.

Figure 45.2 – Deflationary and inflationary gap

Inflationary (expansionary) gap

If, however, at Y_{FE} households expect property prices to rise and the income effect of this causes AD to increase (AD_0 to AD_1 in Figure 45.2) the effect would be inflationary. Since the economy is operating at full employment and no more available factors exist, there is no increase in real output and an inflationary gap is created.

> **Definition: 'Inflationary' and 'expansionary' gap**
>
> When de facto output in terms of macro equilibrium is above potential output, there is a positive output gap – an **inflationary gap**.

It should be noted that the Y_{FE} level of output is pretty much out there in mermaid territory – e.g. it is a guesstimate at best! The OECD collects data on a monthly basis and estimates the positive (inflationary) and negative (deflationary) output gaps:

	Australia	Austria	Canada	Germany	Greece	Iceland
2001	-0.4	0.6	1.1	0.6	-1.5	1.7
2002	0.2	-0.3	1.2	-0.7	-2	-1.5
2003	0.4	-1.9	0.4	-2.2	0	-2.1
2004	0.5	-1.7	0.9	-2.5	0.6	1.9
2005	1	-1.1	1.3	-2.6	-0.2	5.6
2006	0.2	0.6	1.5	0.2	2.5	4.6
2007	1.5	2.4	1.3	2.2	2.6	6.2
2008	0.5	1.6	-0.1	1.6	-0.1	3.3
2009	-1.2	-3.7	-4.4	-4.5	-4.6	-3.9
2010	-1.8	-3	-2.8	-2.3	-8.9	-7.7
2011	-2.6	-1.8	-2.4	-0.8	N.A.	-6.2

(Source: OECD Economic Outlook 90 Database)

The onslaught of the recession starting in 2007/'08 is clearly visible in this selection of economies, where all six saw decreases in aggregate demand during 2009. Australia seemed to weather the storm reasonably well and Germany bounced back the quickest. Which country seems to have been hardest hit?!

Effects of an increase in AD at low and high levels of income

An increase in aggregate demand at very low levels of income (recessionary gap) might not create inflationary pressure at all since there is an abundance of factors, excess capacity and firms have unsold stocks. This is shown in Figure 45.3 where aggregate demand increases from AD_0 to AD_1 and there is an increase in real GDP but no increase in the price level. Accordingly, when the economy is at the full employment level, any increase in aggregate demand (AD_2 to AD_3) leads solely to inflation and no increase in output.
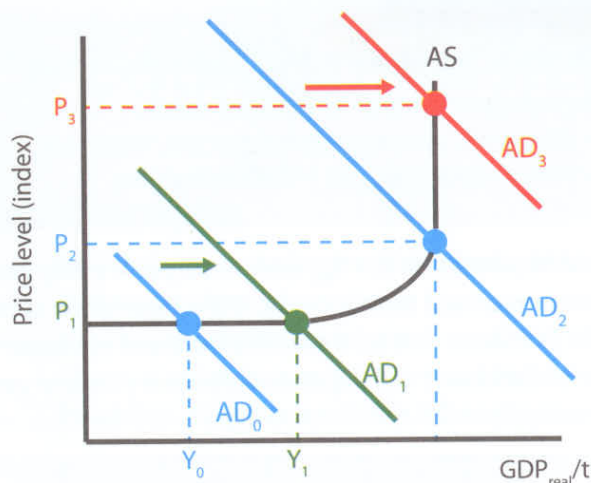
Figure 45.3 – An increase in AD below and at full employment

Possibility of equilibrium below full employment in the long run

The Keynesian model assumes that high unemployment, excess capacity in firms, high levels of stocks and wage stickiness all contribute to perfectly elastic supply at low levels of income. Furthermore, markets are inherently unstable and do not necessarily clear immediately. A key element in Keynesian theory is precisely that markets are *imperfect*, e.g. labour and goods markets do not necessarily clear in the short run. There are built-in market imperfections such as stickiness in labour prices, unions and general unwillingness of workers to accept lower wages in recessions.

Thus, while the basic equilibrium output of AS=AD is attained at below full employment levels, other markets – notably the labour market – need not be in equilibrium. The view is therefore that equilibrium can occur and be maintained at below the full employment level even in the long run – resulting in the prescription that government intervention on the demand side is necessary for full employment to be attained. If governments do not intervene to move the economy towards full employment, it is quite possible for the economy to remain at Y_0 for an indeterminate period. This is a pivotal argument for Keynesian economists in the Keynesian – new-classical debate.

Summary & revision

1. Equilibrium occurs when AD equals AS – e.g. planned expenditure equals planned output.

2. **Output gaps** show the 'distance' between de fact (actual) output and potential output.

   a. When de facto output is below potential (LR) output there is a negative output gap – a **deflationary gap**.

   b. When de facto output is above potential (LR) output there is a positive output gap – and **inflationary gap**.

3. An increase in AD can have different effects on inflation and income in the Keynesian model:

   a. At low levels of income (recession) AD can increase income *without creating inflationary pressure* since there is factor abundance, excess capacity in firms and high levels of stocks.

   b. **Approaching the full employment level** of output, there is a 'trade-off zone' where an increase in AD *increases income* and also creates **inflationary pressure**.

   c. **At full employment** it is impossible to increase output – any increase in AD is **purely inflationary**.

4. The Keynesian model posits that **markets are inherently imperfect** and will not clear in the short run. It is therefore possible to have **macroeconomic equilibrium at below the full employment level** of output.

# 46. Equilibrium in the Monetarist/New Classical Model

Key concepts:

- New classical view of macroeconomic equilibrium in the short run
- New classical view of macroeconomic equilibrium in the long run
- Business cycle revisited using the new-classical model

There is a degree of broad consensus on the *short run* aggregate supply curve, in that it is upward-sloping. Just as in the simple supply and demand model, equilibrium occurs when aggregate supply equals aggregate demand, i.e. when planned expenditure (quantity of real GDP demanded) equals planned output (quantity of real GDP supplied) in an economy. However, the consensus view is limited to the short run, since there is still stark disagreement between the two broad schools of thought – Keynesian and new-classical – as to the shape of the long run aggregate supply curve.

The diagrams in Figure 46.1 illustrate the issue. The left diagram shows the broad consensus on the 'middle range' along a short run aggregate supply curve. When quantity demanded of real GDP equals quantity supplied then macroeconomic equilibrium is attained. This occurs when real GDP is Y_{FE} and the price level is P_0.
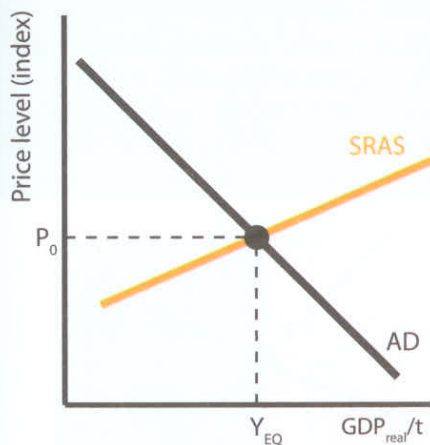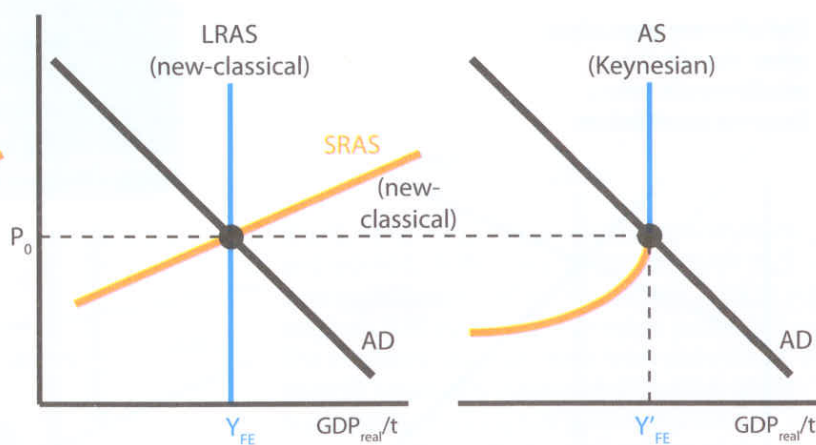
Figure 46.1 Consensus and continued disagreement

New-classical view of short run macroeconomic equilibrium

We shall look at two scenarios, the first when the short run equilibrium is *above* potential output levels and the full employment level of real GDP. Figure 46.2 shows such a situation where output is at Y_1 – which is above the full-employment equilibrium at Y_{FE}. Since real GDP exceeds potential GDP, there is inflationary pressure created by this gap between de facto and potential output. The distance between Y_{FE} and Y_1 is the *inflationary gap* in the economy.

Inflatory gap: arises when the short run equilibrium is above long-run equilibrium.
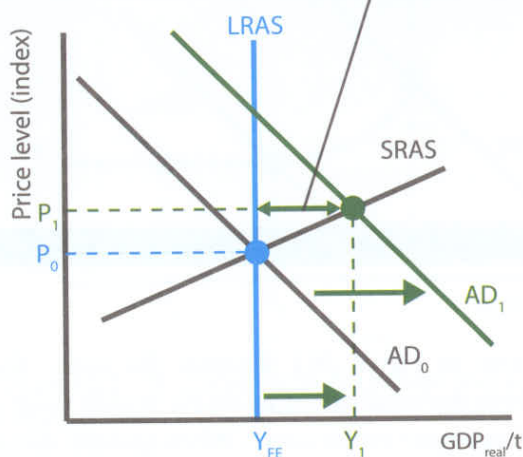
Figure 46.2 Short run equilibrium and inflationary gap in new-classical model

While nothing has been built into this scenario outlining *why* aggregate demand increased, the question is really what happens *after* this point? This is a very central issue in macroeconomics and deals once again with the theoretical rift between Keynesian and new-classical economists.

Deflationary gap: arises when the short run equilibrium is below long-run **equilibrium.**
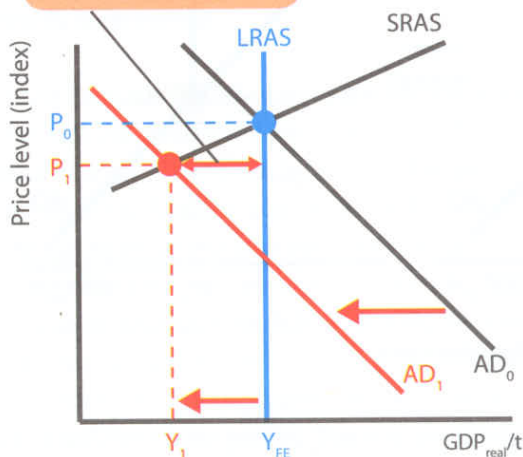
Figure 46.3 Short run equilibrium and deflationary gap in new-classical model

When output is below the long run output level there will be downward pressure on the price level, i.e. *deflationary* pressure. Figure 46.3 shows how the short run equilibrium of Y_1 is below the full employment level, Y_{FE}, and thus potential long run aggregate supply. This is the **deflationary gap** (also

known as the **recessionary gap**) and shows that the economy is either experiencing a recession (see *Business Cycle*) or that the economy's long run potential is outstripping (= exceeding) the short run increase in real GDP.

Referring to an inflationary/deflationary gap is a bit like going through simple supply and demand curves one at a time. It doesn't really mean anything until put into context, i.e. utilised in an economic analysis as an illustration. The most important thing for now is that you understand the central issue in using the new-classical model, namely that economic forces arising during short run equilibrium will assert themselves and seek to restore long run equilibrium at Y_{FE}.

New-classical view of *long run* macroeconomic equilibrium

The long run aggregate supply curve has thus far been explained as the level of potential output in an economy in the long run. This output level is known as **general equilibrium** – all factor markets and goods markets have cleared. It is also the *full employment* level of output since all available labourers willing and able to accept jobs are employed at the going wage rate. Figure 46.4 illustrates the long run macroeconomic equilibrium.

When the economy is operating **above potential output**, there will be **overfull employment**. As real wages rise, labour markets adjust, and both factor and goods markets will clear, restoring the unemployment level to long run equilibrium. The economy returns to its long run equilibrium, Y_{FE}.
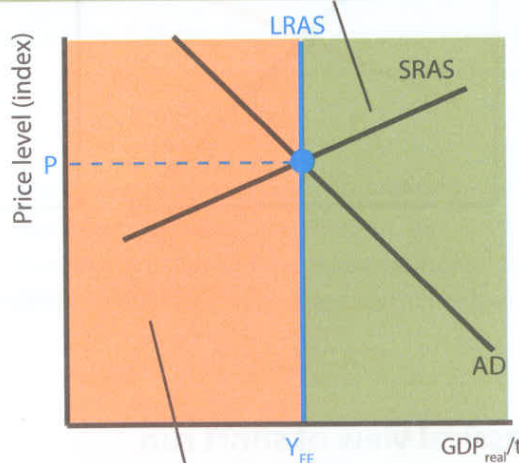
When the economy is operating **below potential output**, there will be **additional unemployment**. In the long run, factor prices will adjust, real wages will fall, and both factor and goods markets will clear. The economy will again be operating at long run (general) equilibrium, Y_{FE}.

Figure 46.4 – LRAS and potential output

The point of intersection between aggregate demand and short run aggregate supply also intersects with long run aggregate supply. At this point output equals the full employment level. **Full employment** is defined as the level of unemployment in an economy existing when supply of labour equals demand for labour, i.e. when the labour market has cleared. As the market for labour is intimately connected to output the demand for labour will be derived from planned output and planned expenditure. Those without jobs at full employment level on the labour market are 'between jobs', and since there will always be a portion of the total available work force seeking employment, full employment renders a *'natural' rate of unemployment* (see Chapters 50 and 54).[1]

Strong economic forces will be enacted in a situation where the economy is operating above or below the full employment level of output of Y_{FE} in Figure 46.4. As the new-classical AS-AD model assumes that labour markets function perfectly, any disequilibrium in the labour market will be short term so when markets clear the economy will be at full employment no matter what the price level in the economy is. Keep in mind that the

1   One question begging to be addressed is how we arrive at the point we call potential output in the first place. Good question. I don't have an answer. Unfortunately, nobody else does either. In fact, it is impossible to precisely set down potential output in an economy and the LRAS curve must be viewed as a theoretical assumption put in place to aid us in our analysis. So while there is theoretical merit in *assuming* that potential output is unrelated to the price level and thus vertical, the actual position of long run aggregate supply must be *estimated*.

new-classical view is that labour markets function perfectly so any disequilibrium (unemployment) in the labour market will be short term only.

One of the 'strong economic forces' referred to above is of course **inflationary pressure**. Inflation is rather carefully defined in economics, and means a *sustained rise in the general price level*. Since the macroeconomic environment is most dynamic over time, price levels will be continually adjusting to economic activity. In fact, 'long run' in our AS-AD model doesn't really mean 'the point where the economy will wind up ultimately' but rather 'the level around which economic activity will hover over the longer term'. In other words, it is highly unlikely that the economy will be at general equilibrium for any length of time since there is always an ongoing process of adjustment *towards* the long run output level – not *at* the long run output level.

Coming up next; a *Most Important Diagram*, as my upcoming bestseller *Economics According to Winnie the Pooh* will read. Central to the new-classical view is that demand management is relatively ineffective in the long run. Figure 46.5 shows the long run effect of expansionary demand management – increasing government spending and/or lowering interest rates – with the economy initially at full employment level of output, **point A**.

MACROECONOMICS
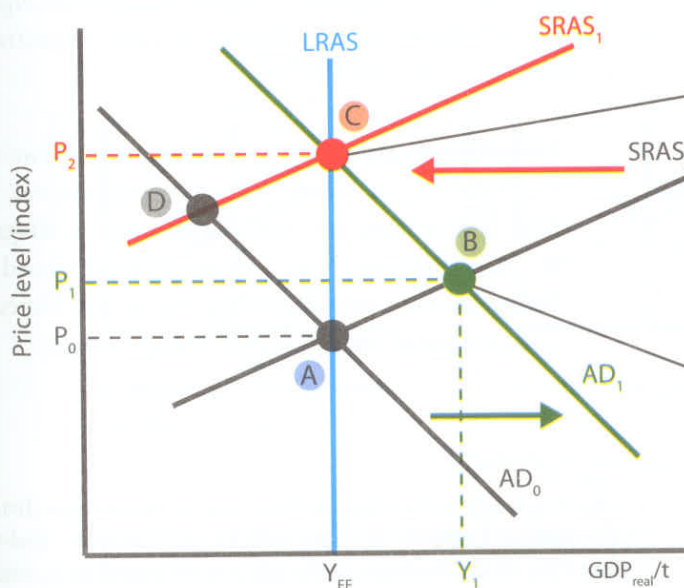
B to C: As factor prices are bid up by both firms and workers, in the long run the short run aggregate supply will shift to the left, back to the original LRAS curve but at a higher price level. The economy will thus move from point **A** to **C** via point **B**. Again, any increase in AD which is not matched by the long run potential of the economy will be inflationary – shown by the increase of the price level from P_0 to P_2.

A to B: An increase in AD will increase output in the short run, but also make factors scarce and decrease real wages for workers.

Figure 46.5 – New-classical view of an increase in aggregate demand beyond full employment

A to B: The increase in aggregate demand due to fiscal/monetary stimulation increases aggregate demand from AD_0 to AD_1; real income increases from Y_{FE} to Y_1 and the price level rises from P_0 to P_1. Since wage levels and factor prices are assumed to remain constant during the short run, the short run supply curve does not shift. Instead firms experience higher costs due to bottlenecks and scarcer factors.

B to C: The short run effect is an inflationary gap at **point B**. In line with derived demand effects on factors, this will result in ever scarcer factors of production – e.g. capital, raw material and labour – and the increase in the price level from P_0 to P_1 will lead to a hollowing-out of wages for workers. Factor prices will ultimately be forced upwards as firms bid on available factors and workers will demand higher wages to make up for lost real purchasing power; short run aggregate supply will decrease in the long run, shown by the shift from $SRAS_0$ to $SRAS_1$. The economy has returned to the full employment level of output, Y_{FE}, but at a higher price level, P_2. This is **point C** in the diagram.

The A-B-C series in Figure 46.5 once again shows the importance of the long run aggregate supply in the new-classical AS-AD model. Any point of short run equilibrium below or above the long run potential – shown by the LRAS curve in the model – will ultimately be corrected as factor markets clear.

A fall in AD – point C to D to A: I briefly illustrate the new-classical view of the long run by assuming that the economy is initially in equilibrium at **point C**, and that aggregate demand instead has decreased, shifting aggregate demand from AD_1 to AD_0 moving the economy to **point D** and a resulting deflationary gap. As firms lower prices in accordance with falling demand, the gap between (falling) final output prices and wage prices narrows, this in fact means higher relative labour prices for firms. Firms respond by demanding less labour and as the labour market responds, labour prices fall – which enables firms to increase output and increase short run aggregate supply. This increases the short run aggregate supply from $SRAS_1$ to $SRAS_0$ bringing the economy back to general long run equilibrium at **point A**.

Business cycle revisited using the new-classical model

> "Economists have forecast 9 out of the last 5 recessions." Unknown.

The observation that economic activity varies over time is not a new discovery. I always find it fitting irony that in many Germanic languages the word for cyclical variations in the business cycle is *Konjunktur*, derived from the Roman word *conjugare* – 'to bend, bring together, align' – which was used in Roman times to explain planetary alignments in astrology. Rulers and the ruled have for centuries known that general prosperity fluctuated between 'good times' and 'bad times' and there have always been attempts to explain and predict the economic future. Perhaps the most interesting of these, mentioned earlier, was a proposal by William Jevons (one of the original neo-classical economists from the late 1800s) that the fluctuations in the economy were caused by sunspots.[2]

While economists today seldom use planetary alignments (or the entrails of geese and dogs) to predict business activity, I am not entirely certain that we are much better at predictions than the Roman fortune-tellers and astrologers – as economists were then called. Economists today have as yet to construct a method for accurately predicting the sequences of economic expansion and contraction, but not for want of trying. What we refer to as the business cycles (or trade cycle) is the *aggregate of economic activity* over a longish period of time, and economists use a number of variables in trying to map out the changes. The most common variable used and correlated to time is real GDP. Business cycle theory then looks for the causes of business cycles and of possible responses by policymakers. We will look at several possibilities herein, but first let's examine the basic cyclical variation through the lens of the AS-AD model more closely.

Aggregate supply/demand and the business cycle

The diagram series in Figure 46.6 is a continuation of Figures 46.2 and 46.3 (inflationary and deflationary gap, respectively) covered earlier.

- **Deflationary gap:** Point A in diagram I shows that real GDP, Y_0, is below the long run potential, Y_{FE}. This is illustrated as a deflationary gap in **diagram II**. A deflationary gap is often characterised by falling inflation rates (or falling prices) and increased unemployment (called *cyclical* unemployment).

---

2   This is actually not as zany as it may sound. Jevons saw distinct correlation between the periodicity of sunspots and the duration of the business cycle. He hypothesized that sunspot activity caused changes in the earth's weather cycle, and that this then affected crops and thus the entire agrarian economy. Unfortunately it was discovered that the calculations of sunspot activity were in error and their length did not actually match the business cycle. Back to the drawing board.

- **General equilibrium:** Over the next time period (note that t_0 is a *point* in time and t_0 to t_1 is a *period* of time) aggregate demand increases, bringing short run AD and AS in line with long run potential at Y_{FE} at point B in diagram I. As aggregate demand rises between t_0 and t_1, unemployment decreases and settles at the full employment level and real GDP increases from Y_0 to Y_1/Y_{FE}. General equilibrium is achieved (AD = SRAS = LRAS) as seen in **diagram III**.

- **Inflationary gap:** Aggregate demand continues to increase during the next period, (t_1 to t_2), resulting in real GDP of Y_2 at point C and an inflationary gap shown in **diagram IV**. Unemployment decreases below the full employment level. Inflationary gaps are associated with, well, inflationary pressure.

- **Points D and E** are real output levels when the AD trend is reversed, in other words when aggregate demand is falling over the time period t_2 to t_4.

In moving from point A to point E in Figure 46.6 a cycle is completed. Please note that the series is only an *illustration* and a rather simplistic one at that. I am trying to explain the basic business cycle – not reality! I have in this respect made two highly simplifying assumptions; 1) real potential GDP has not changed over the cycle; and 2) only aggregate demand changes during the cyclical phases which is rather unrealistic. If you can answer Pop Quiz 3.3.4 question number one below correctly then you are well on your way to understanding this.
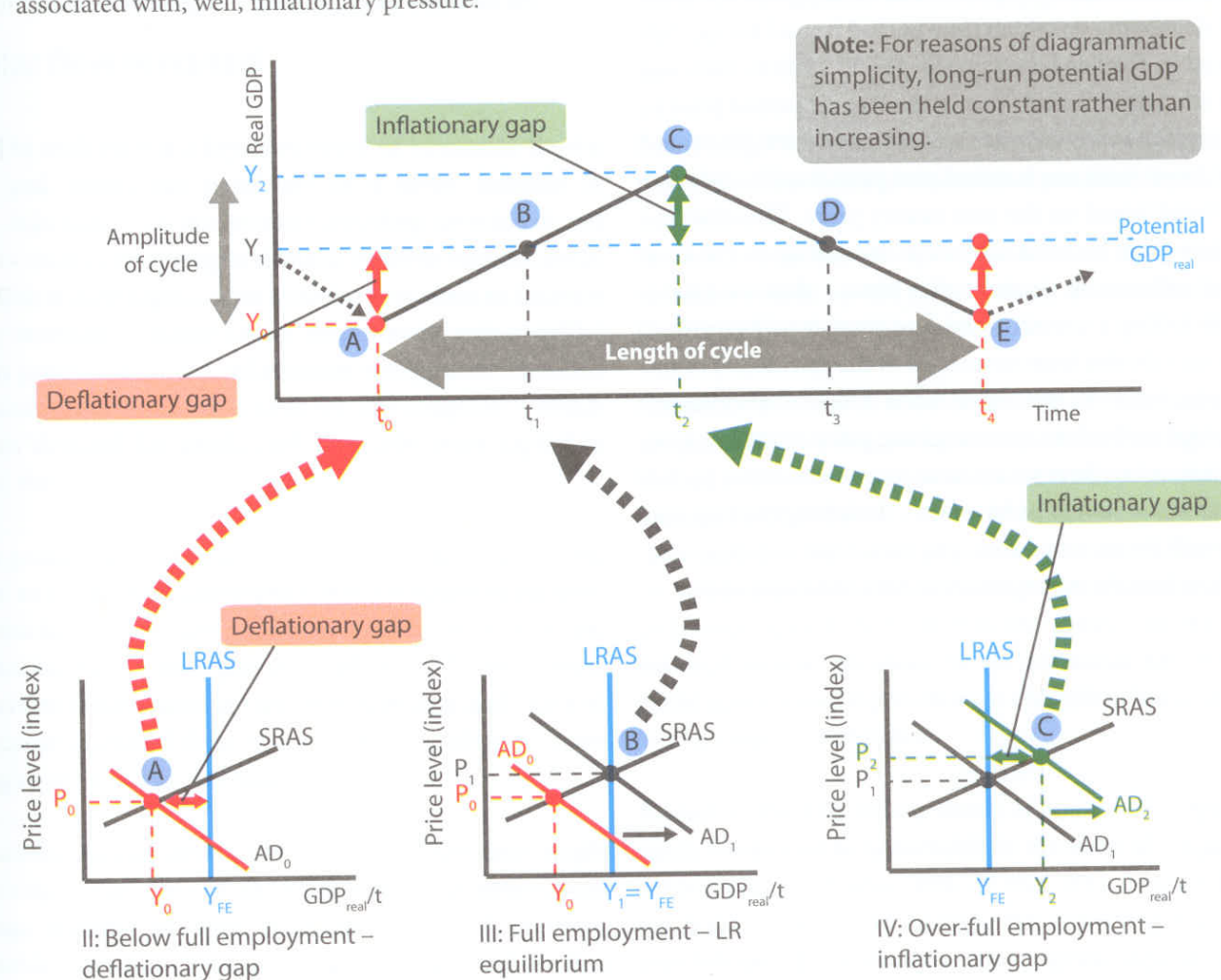

Figure 46.6 – AD and the business cycle

Based on twenty years of real GDP values for Sweden, where the trend line has been adapted in order to remove cyclical variations, basically, the sum of the positive and negative output gaps should be zero. Over the 20 year period, real GDP in Sweden increased from SEK1,350 billion to SEK1,900 billion. This shows that long run potential output during the time period has grown by some 40%, and, solving the equation (X = average growth rate); $1{,}350 \times X^{20}1{,}900$ gives a yearly average increase of real GDP of 1.723%. This is significantly lower than the average growth rate of 2.75% for other developed countries during the same time period.[3]
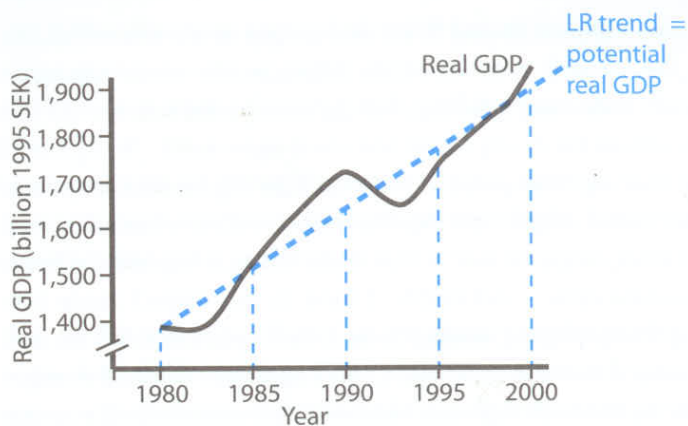
3   UNCTAD handbook of statistics 2002, p 347

Figure 46.7 Cyclical variations and long run growth in Sweden 1980 – 2000

The trend method used above comes with a most noticeable weakness attached, namely that the trend for any given economy will differ depending on **which time period** is used to calculate it. The trend in Sweden based on the period 1980 to 2000 was 1.7%, but other figures show that in the 50 year period prior to 2000 the average yearly increase was closer to 2.8%. The trend over a five, ten or even one hundred year period will be different from the trend based on the past twenty years. Therefore the long run trend can be most subjective depending on the time period and will not be constant over time – thus we have a problem in setting a correct trend line since it will vary over the longer and shorter time series used in the calculation of the trend. In conclusion to the above, while I have not addressed any additional methods in assessing long run potential output, the fact remains that since we cannot agree on what time period the long run trend should be based on – 'how long the long run is' – there will be no conclusive way to set the waterline and agree on how high the waves and how deep the swells are.

Summary & revision

1. There is **broad consensus** amongst economists that the *SRAS curve is upward sloping*. The main disagreement is LRAS.

2. The **new-classical view** is that the *LRAS curve is not correlated to the price level*. It is possible for AD to create short-run equilibrium beyond income at full employment (LRAS) but rising factor costs and demands for higher wages will push up production costs for firms and decrease SRAS back to general equilibrium.

3. The LRAS curve shows **potential long run output** and is an estimate based on long run growth trends.

47. The Keynesian Multiplier

Key concepts: HL extensions

- Circular flow revisited
- Calculation of the multiplier (k)
- Diagrammatical illustration
- Evaluation of the multiplier

Q: How many Keynesian economists does it take to change a light bulb?
A: All. Because then you will generate employment, more consumption, shifting aggregate demand …

Circular flow revisited

The multiplier is a key component of Keynesian theory, and shows the possibility of a given increase in injections – i.e. government spending, investment and exports – increasing aggregate demand by more than the initial value. This is quite logical at an intuitive level since an increase in, say, investment might create employment opportunities in firms producing capital, whereupon newly hired labourers will receive income which is used for consumption – which increases demand for goods and ultimately more capital to produce the goods.

Keynes posited that the main influence on consumption was income; or rather that **consumption was a function of income**. Given that households have a **marginal propensity to consume**, an initial increase in aggregate demand caused by an increase in injections (investment, government spending and exports) would cause increased flows in the economy leading to larger final aggregate demand and national income.

I will use the circular flow model to illustrate how there might be a multiplicative effect of an initial increase in government spending. Assume an open economy comprised of firms, households, financial institutions, government and foreign sector. An increase in any of the aggregate demand components will mean a larger flow around the system, in other words, an increase in aggregate demand. The theory of the multiplier states that an increase in injections of $100 will cause final national income to increase by *more* than $100 – depending on the value of the multiplier. The question here is *how much* aggregate demand will ultimately increase due to a change in injections, i.e. what is the value of the multiplier.

Posit that government wishes to implement expansionary fiscal policies and increases government expenditure by €50 billion.

Round 1: This initial expenditure, round 1 in Figure 47.1, increases output by €50 billion which flows to households in the form of wages/rent/interest/profit.

Round 2: In round 2, households will use this income in two ways; a portion will go to *taxes*, *imports* and *savings* – this is a leakage out of the system – and a portion will go to consumption. Assume that €20 billion – 40% – leaks out of the circular flow in this manner and that the rest goes to consumption expenditure, i.e. €30 billion (€50 - €20 in leakages) of households' increase in income goes right back into the system. In this manner, national income increases by an *additional €30 billion*. The initial government expenditure of €50 billion has now caused a total increase of €80 billion.

Round 3: The €30 billion worth of additional output is once again transferred to households in the form of wages/interest/rents/profits, and after 40% of this (€12 billion) leaks out via taxes, imports and savings, the remaining €18 billion is injected into the economy as consumption expenditure in the third round. National income will now have increased by €98 billion as each successive round adds more to the flow in the system. Note, however, that each consecutive round adds ever less to national income; this is the leakage effect where 40% of each new round is going to taxes, imports and savings and is therefore withdrawn from the system.
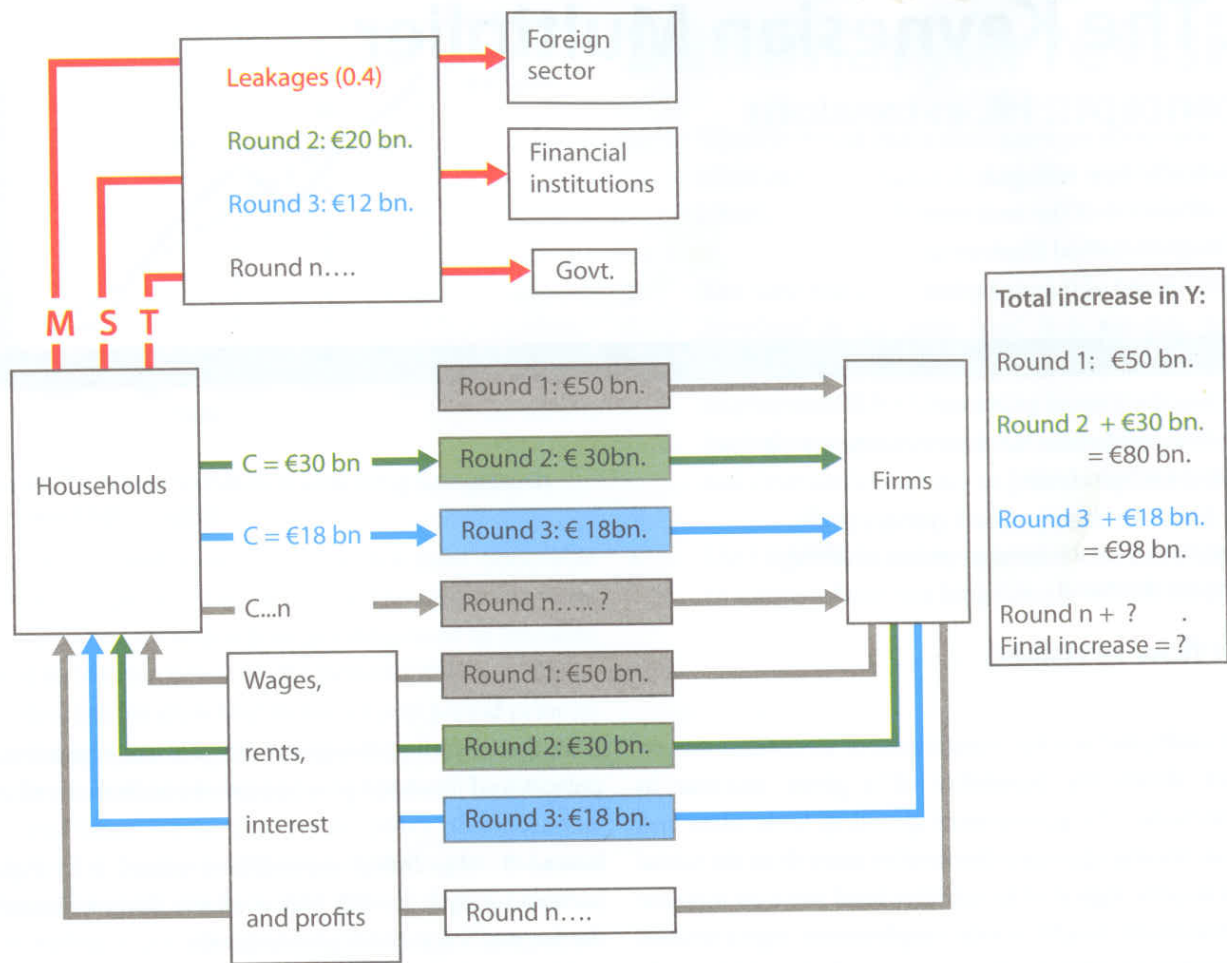
Round n: How much will the final value be?!

Figure 47.1 Circular flow and the Keynesian multiplier

Sparing you some lengthy and tedious math (I think my editor mentioned something about 'infinite geometric series') the final value of all output resulting from the initial government injection of €50 billion will be €125 billion. The multiplicative effect shows how much final expenditure will be as a result of an initial increase in injections. This gives us a ratio, $\Delta Y/\Delta J$, which is the *value of the multiplier*. In the example here, when the government primed the economy by increasing government spending by €50 billion and aggregate demand as a result ultimately increases by €125 billion, then the initial injection has **multiplied by a factor of 2.5**, i.e. 125/50 billion.

Calculation of the multiplier (k)

The value of €125 billion in the previous example was not pulled from a hat, but calculated using the formula for the multiplier – which is actually fairly straightforward, but let's go through it one step at a time.

Marginal propensity to consume (MPC)

In the Keynesian model, households' **marginal propensity to consume**, MPC, is the proportion of any increase in income used for domestic consumption. The MPC is the change in consumption over the change in income; $\Delta C/\Delta Y$.

This proportion of increased consumption in the example is 0.6 (i.e. 60% of any increase in household income goes to consumption), which is the *marginal propensity to consume*. In the first round in Figure 47.1, the increase in (Y) of €50 billion in government spending resulted in an increase in consumption (C) of €30 billion.

$$MPC = \frac{\Delta C}{\Delta Y} = \frac{30 \text{ billion}}{50 \text{ billion}} = 0.6$$

Marginal propensity to leak (MPL)

Correspondingly, the 0.4 proportion which does not go to consumption is leaking out of the system; *savings, taxes* and *imports*. This proportion of a change in income which does not return to the system in the form of consumption is the **marginal**

propensity to leak, MPL, which is comprised of the marginal propensity to **save** (MPS), the marginal propensity to **import** (MPM) and the marginal propensity to **tax** (MPT).[1]

$$MPS = \frac{\Delta S}{\Delta Y} \quad MPM = \frac{\Delta M}{\Delta Y} \quad MPT = \frac{\Delta T}{\Delta Y}$$

Let us assume that out of the €50 billion increase in income, households save €2.5 billion (5% or 0.05), use €7.5 billion (15% or 0.15) for imports and have to pay €10 billion in tax (20%).

$$MPS = \frac{\Delta 2.5}{\Delta 50} \quad MPM = \frac{\Delta 7.5}{\Delta 50} \quad MPT = \frac{\Delta 10}{\Delta 50}$$

$$MPL = 0.05 + 0.15 + 0.2 \rightarrow MPL = 0.4$$

$$MPL = \frac{\Delta S + \Delta M + \Delta T}{\Delta Y} \rightarrow = \frac{20 \text{ billion}}{50 \text{ billion}} = 0.4$$

Summing up so far:

- Households' increase in consumption due to an increase in income is the marginal propensity to consume; $MPC = \Delta C / \Delta Y$. In the example given the **MPC is 0.6**.

- The remaining 0.4 is a leakage out of the system. The marginal propensity to leak is comprised of saving (MPS), import expenditure (MPM) and taxes (MPT). The sum of leakage propensities is MPS + MPM + MPM. In the example, the **MPL is 0.4**.

- Basically, any increase in income can go only two ways; consumption or savings/imports/taxes. Thus, the sum of the marginal propensity to consume and the marginal propensity to leak equals 1. **MPC + MPL = 1.**

Keynesian multiplier (k)

Finally, we arrive at how to calculate the actual multiplier, 'k' (for 'Keynesian' naturally). The initial injection of €50 billion in government spending will reverberate within the flow, with each successive round giving a smaller and smaller increase in consumption; €50 billion in the first round, €30 billion in the second, and so on. The *final* increase in national income is given by: *initial change in income times k*.

1 You don't need to be Werner von Braun to realise that the sum of the marginal propensity to consume and the marginal propensity to leak must always equal one. This makes perfect sense, since any increase in household income will either be used for domestic consumption or savings/taxes/imports. Thus; MPC + MPL = 1, and 1 − MPC = MPL.

The formula for the Keynesian multiplier is;

$$k = \frac{1}{MPL} \text{ or } k = \frac{1}{(1 - MPC)}.$$

Yes, the denominator is of course the same in both versions, e.g. the MPL = 1 − MPC.

Version 1; $k = \frac{1}{MPL} \rightarrow 1/0.4 = 2.5$

Version 2; $k = \frac{1}{(1 - MPC)} \rightarrow \frac{1}{1 - 0.6} = 2.5$

$\Delta \text{initial J} \times k = \Delta \text{final Y}$

Thus; €50 billion x 2.5 = €125 billion

'J'; injection (here, government spending)

'k'; the value of the multiplier

'MPC'; marginal propensity to consume $\frac{\Delta C}{\Delta Y}$

'MPL'; marginal propensity to leak $\frac{\Delta L}{\Delta Y}$

MACROECONOMICS

The initial increase of €50 billion in government spending led to an ultimate increase in national income of €125 billion, giving a multiplier of 2.5. This value of k is very high and depends on the low level of leakages – 0.4 – out of the system during each round. If we instead apply a more realistic leakage component, say a total leakage of 0.8, then the multiplicative effect (k) will be 1/0.8, i.e. a value of 1.25 rather than 2.5. This is in keeping with the formulae above; **the higher the leakages** (taxes, imports and saving) **the lower the multiplicative effect** of any given increase in government, investment or export expenditure. This is the same as saying that the higher the marginal propensity of households to consume, the higher the value of the multiplier. By affecting the marginal propensity to withdraw (leak), government can enhance the multiplicative effects in the economy: lower interest rates will lower the MPS; lower income taxes will lower the MPT; and barriers to trade will lower the MPM. Any and all of these actions would theoretically lower the MPL – *ceteris paribus* – and thus increase the value of the multiplier, k, which would increase the effects of policies aimed at influencing aggregate demand.

Diagrammatical illustration

It probably makes intuitive sense that at very low levels of income the impact of the multiplier will be greater, since pent-up demand amongst labourers and households will lead them to spend a the greater proportion of an increase in incomes. In

Figure 47.2 below, the increase in government spending shifts AD from AD_0 to AD_1 and the full force of the multiplier comes into play, e.g. final output increases from 100 to 250.[2]

The **multiplicative effect on AD is zero** when all factors are occupied and the economy is operating at the full employment level of output.

The **multiplicative effect is lessened** as the economy approaches full employment.
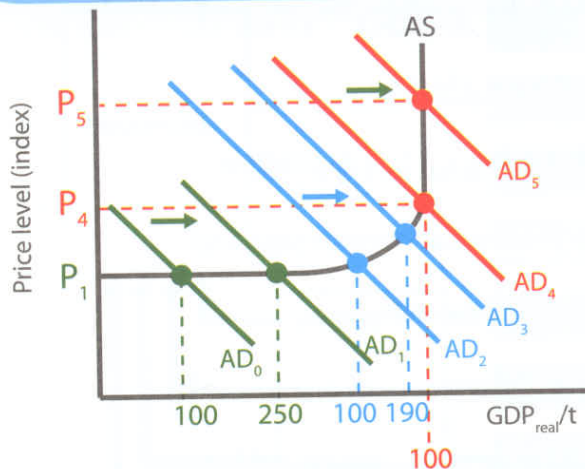
At low levels of income the **full value of the multiplier** (2.5) feeds through the economy.

Figure 47.2 The Keynesian multiplier at different levels of income

Evaluation of the multiplier

There is an obvious appeal of the multiplier in demand-management since a government will theoretically be able to calculate a larger final increase in GDP than initial government spending. In fact, governments have counted on multiplicative effects in utilising demand-side fiscal policies in order to achieve balanced budgets – in spite of deficit spending – over a business cycle. The multiplicative increase in income has been factored-in when deficit spending and/or reduced taxation has been undertaken. This brings us to an additional reason for the reduced influence and therefore use of fiscal policies during the 1970s and '80s, namely the **increased openness** and ease of trade between countries. Lower barriers to trade such as tariffs and quotas, plus deregulation of financial markets increased the ability of households to consume goods from abroad. Increasingly goods and services could be purchased via import,

which meant that any given increase in stimulatory government spending would be withdrawn from the domestic flow of economic activity. The *increase in the marginal propensity to import* – i.e. leakage – meant a corresponding fall in the value of the multiplier, hence a decrease in the effectiveness of fiscal policies aimed at creating a multiplicative effect by way of increasing disposable income.

Another obvious line of criticism is the simple fact that governments need to raise money in order to increase government spending. There are some serious trade-off effects here:

- Increased **government debt** is a burden on future households and might lead to higher taxes.

- Deficits and debt both have consequences on the **exchange rate** in the long run. Higher debt ratios (as a percentage of GDP) are clearly linked to weaker currencies. (See Chapter 67.)

- If the government borrows money on the open market there will be an effect on the rate of interest – the cost of borrowing money rises in the economy. This might cause a '**crowding out**' of private sector investment as firms scale back investment plans due to higher costs of funding loans. The decrease in investment will have a dampening effect on aggregate demand. (For full explanation see Chapter 57.)

---

2 I have indexed the starting values for the sake of illustrative simplicity. I have also assumed that taxes and imports are zero!

POP QUIZ

Keynesian multiplier

1. In an economy, government spending increases by $10 billion and final income increases by $25 billion. What is the value of the MPC?

2. If government spending increases by €10 billion and the *marginal propensity to consume* (MPC) is 0.5, what will the final increase in national output be?

3. How will an increase in the *marginal propensity to save* (MPS) affect the multiplier?

4. Looking ahead: How might the creation of a common market (see Chapter 74) affect individual countries' ability to rely on multiplicative effects in using fiscal policies to stimulate the domestic economy?

5. Evaluate the extent to which an increase in government spending might have multiplicative effects – e.g. what might strengthen or weaken the impact on final output?

Summary & revision

1. In the **circular flow model**, *injections* (J) are I, G and X. *Leakages* (L) are S, T and M.

2. Government spending is a component of AD. The Keynesian multiplier shows how an increase in government spending results in a final increase in GDP that is greater than the initial amount of government spending.

3. The MPC is the change in consumption due to a change in income; $\Delta C/\Delta Y$.

4. The MPL is the change in leakages due to a change in income; $\Delta L/\Delta Y$. The MPL is comprised of:

   a. The marginal propensity to save (MPS)

   b. The marginal propensity to import (MPM)

   c. The marginal propensity to tax (MPT)

   d. Hence, the MPL is the sum of MPS, MPM and MPT

5. The **Keynesian multiplier** (k) is given by:

   a. 1/MPL, or

   b. 1/1-MPC

6. MPC + MPL = 1. If households spend 60% of any increase in income then the MPC is 0.6 and the MPL is 0.4.

7. The multiplier works better at low levels of income than at high levels.

8. **Weaknesses** in using the multiplier include:

   a. Increased **openness in trade** increases the MPM

   b. **Marginal tax levels** in many countries have increased the MPT

   c. Governments might finance spending via **increased deficits and debt** which can have adverse effects on taxes, inflation and exchange rates

   d. Government borrowing might lead to higher interest rates and 'crowd out' private sector investment – which is a component of AD.

2.3

48. Overview of main macro objectives – "Big 5 + 4"

Key concepts:

- Main macro objectives
- Other macro goals
- Back to trade-offs

Main macro objectives

Section 2.3 (Chapters 48 to 55) looks at the macroeconomic goals that economies strive towards and Sections 2.4 to 2.6 (Chapters 56 to 62) deals with the methods available to attain them. As outlined in Chapter 36, the main goals are:

1. **High growth** – measured by the increase in real GDP or GNP over a period of time

2. **Low unemployment** – measured by the percentage of the total labour force not holding a job

3. **Price stability** – measured by the consumer price index (CPI)

4. **External equilibrium** – measured by exchange rate equilibrium and the balance between imports, exports and capital flows in and out of the economy

5. **Equity in the distribution of income** – measured by the Lorenz curve with the Gini coefficient and the Kuznets ratio

Other macro goals

While no less important (though I can be accused of being rather normative now!) there are additional objectives which countries seek to fulfil:

A. **Sustainability** in the use of resources – measured by various environmental indexes and indicators

B. **Factor productivity** – measured by output per unit of factor input

C. **Development** – a general increase in the living standards across a broad section of society

D. **Balanced budget** – governments seek to keep a balance between government spending and tax receipts

Back to trade-offs

However tiresomely repetitive this must sound, all the above involve yet more trade-offs in implementation. If there is one thing I wish you to take with you as my student, it is the simple fact that one cannot have it all – everything means giving something up. Our job as economists is to estimate the opportunity costs and point them out.

Do the following: sketch an AS-AD diagram and play around with an economics 'Sim City' scenario where government implements various policies to attain one of the four main macro objectives. For example, government increases government spending to stimulate AD and thus increase growth (1 above) and lower unemployment (2 above). What is the trade-off? Correct; inflation (3 above). And also the possibility that higher inflation leads to lower demand for exports which in turn lowers demand for the home currency. In other words, another trade-off; increased income might lead to an imbalance in trade.

If you have studied the macro chapters diligently thus far you should be able to identify at least five macroeconomic trade-offs. Answers are in the summary.

Summary & revision

1. The **main macro goals** are growth, low unemployment, price stability, external balance and 'fairness' in income distribution.

2. **Other goals** are sustainability, factor productivity and development.

3. Several **trade-offs** often occur in the pursuit of macro goals. Some of the main ones are:

   a. *Growth and price stability* (as an increase in AD can be inflationary)

   b. *Lower unemployment and price stability* (increased AD lowers unemployment as more labour is needed but aforementioned inflationary pressure is present)

   c. *Lower unemployment and a balanced budget* (governments are often forced into deficit spending, i.e. borrow money to increase government spending aimed at reducing unemployment)

   d. *Growth and the balance of trade* (as incomes rise there is a propensity for households to increase consumption – some of these goods are imported and this may cause a trade deficit)

   e. *Growth and the exchange rate* (a government which lowers the interest rate to stimulate consumption and investment – and thus AD – would soon discover that foreign depositors/speculators withdraw funds which would cause the demand for the home currency to fall)

   f. *Growth and the environment* (high growth rates in developing countries are often associated with environmental deterioration)

49. Economic Growth

Key concepts:

- Definition of economic growth
- Shifting LRAS – quantity and quality of factors
- Investment and growth
    - Physical capital
    - Human capital
- Productivity and growth
- Consequences of growth

HL extension:

- Calculating growth from data

Definition of economic growth

An economy 'grows' when the money value of total output adjusted for inflation increases, e.g. there is an increase in real GDP.

> **Definition: 'Economic growth'**
>
> Economic growth takes place when an economy produces more output during a period of time. This is measured in real money terms, i.e. adjusted for inflation. The most common measure is **real GDP change per year.**

Often data will be in sets spanning several years – especially when one wishes to see extraordinary periods such as recessions or booms. Figure 49.1 below shows the effects of the 2007/'08 recession for some of the major economies. It is clear that the recession hit the the BRIC nations (Brazil, Russia, India and China) far less than advanced nation economies such as the US, Japan and Britain.
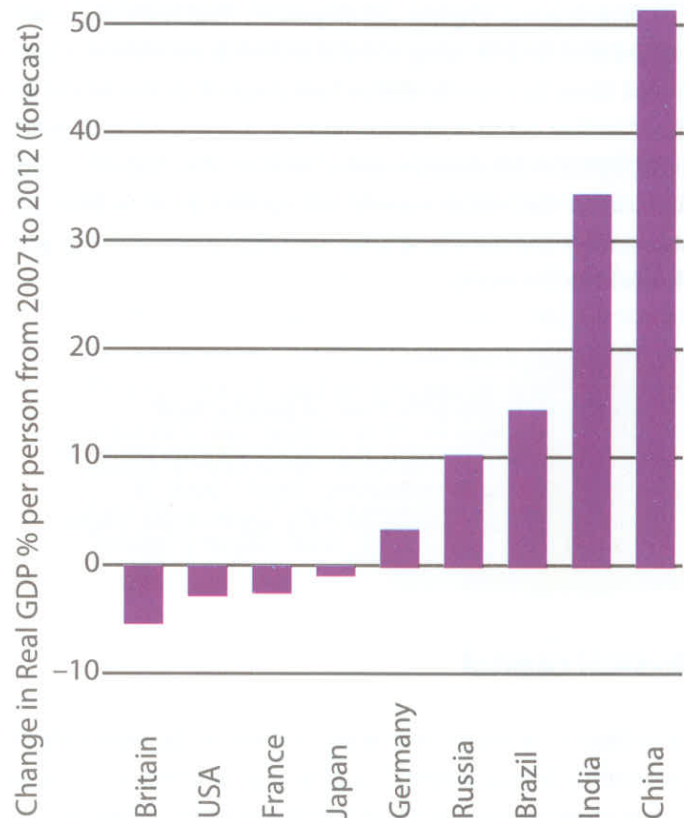

Figure 49.1 The BRIC countries and the 2008 recession as reported by *The Economist*, "*Winners and losers Jan 16th 2012*"

Investment and growth

In the BRIC countries shown in Figure 49.1, GDP per capita is projected to increase far more than employment. This is possible only by way of increasing the productivity of the factors used in production. Three key factors in an economy are land, capital and labour – raw materials, man-made factors of production and people at work using the raw materials and capital. Often economists refer more pointedly to *natural factors*, *physical* capital and *human* capital.

Natural factors

Clearly arable land, abundant forests, plentiful water, and any one of thousands of other land resources are linked to a country's ability to satisfy its citizen's needs and wants. Different countries are differently endowed with such **natural factors**; Canada has the world's largest freshwater supply while Namibia has one of the largest reserves of uranium in the world. Natural factors can to a certain extent explain development; England's abundant coal and iron helped bring about a technical revolution which enabled them to become the richest country in the world by 1900. Another country of some fame in this respect is **Iceland**, whose 320,000 inhabitants enjoy one of the highest living standards in the world; natural hot springs provide plentiful cheap energy and the island is surrounded by one of the largest fish stocks in the world.

> **Definition: 'Natural factors of production'**
> **Natural factors** consist of a country's endowments (= gifts) in natural resources – land – such as minerals, forests, arable land for agriculture, plant and animal diversity … etc.

Physical capital

The definition of investment is an increase in the capital stock of a country during a period of time. Firms invest in order to become more productive and while this is a cost (depreciation) it serves to lower costs and increase profits. Physical capital is *tangible*, meaning 'material' or 'touchable' and refers to the machines, tools, factories and transport vehicles used in the production of goods and services. Man-made factors of production are collectively known as **physical capital**. You will encounter this numerous times in this book as *investment* or an *increase in capital stock*. Machinery and factories are directly involved in production, while roads and communications are necessary 'backbones' of production, e.g. **infrastructure**.

> **Definition: 'Physical capital'**
> An increase in investment means that **physical capital** has increased. Examples are factories, machinery and roads. The accumulation of capital increases a country's capital stock.

Human capital

The combined talents, experience, knowledge and education of labour are major *intangible* assets in firms. Experienced workers will produce more per hour than new and inexperienced workers – just as experienced teachers will be able to get the message across better and with less effort than inexperienced teachers.

Increase in the quality of human resources: The quality of the working population – the **human capital** – is the result of knowledge, skills, education and training. Human capital also results from health care, as this adds to productive capabilities in a country. Most economists would agree that education/ knowledge and health care are central factors in potential growth and development. The benefits of enhancing human resources – *investment* in human resources rather – go far beyond economic growth. A short-list of positive effects resulting from investment in human capital would have both economic and social benefits:

- **Economic benefits** would be higher productivity in the economy; increased labour mobility as more people would be attractive on the job market; more dynamism as a result of entrepreneurial spirit; and better use of finite resources.

- **Social benefits** include better health and longer lives; greater participation and democratisation in local and municipal issues; better opportunities for women in choosing their own lives; and the ability to partake in a wider range of cultural offerings.

> **Definition: 'Human factors' and 'human capital'**
> Human resources – labour – comprise the skills, knowledge, experience, education and health of the population which makes up the labour pool. Investment in these areas, often called social investment, increases **human capital**.

MACROECONOMICS

Productivity and growth

According to the US Department of Labor (yes, it's spelled like that) average productivity increased by 0.4% during 2010 to 2011.[1] Labour productivity is calculated by dividing real output (e.g. the value of output adjusted for inflation) with the total amount of hours worked in all industries in the economy. Naturally this measure also includes the physical capital used in production so better tools, increased use of computerised manufacturing processes and *more* machines will all increase productivity. Out of a very long list, here are some examples of how economies have increased productivity over the past century:

- Spread of new technology and manufacturing techniques between industries and countries.

- Increased division of labour throughout the supply-chain and distribution networks.

- Vastly improved (and cheaper) telecommunication networks which have enabled firms to use skills and expertise in other countries.

- Increasing educational standards and universities linked to research and development centres in private enterprises.

- Greater awareness of how motivational aspects at work increase labour productivity.

## Shifting LRAS – quantity and quality of factors

As outlined above and in Chapter 44, any macro variables which increase the quality and/or quantity of factors of production will shift LRAS. This has historically been going on since … well, since the industrial revolution at the very least. Better tools, more efficient machines, improved education have all served to increase the potential output – LRAS – for hundreds of years. As LRAS increases, the underlying trend rate of growth (see the business cycle in Chapter 46, Figure 46.6) will also increase.

Assume an economy at initial LR equilibrium – also known as general equilibrium or the full employment level of output – where the quantity and/or quality of factors of production increase. The initial shift of LRAS ($LRAS_0$ to $LRAS_1$) in Figure 49.2 gives three possible scenarios over time:

1. **Growth and a degree of inflation:** As the potential economy has moved beyond actual output a negative output gap is created allowing AD to increase from AD_0 to AD_2 creating a new general equilibrium at *point A*. Note that in this new-classical version, the shift in LRAS enables AD to increase without causing wage inflation and thus decreasing SRAS. (Compare with Figure 46.5 in Chapter 46.)

2. **Growth and stable prices:** Any variables affecting LRAS can also affect SRAS. Increasing SRAS from $SRAS_0$ to $SRAS_2$ would have an anti-inflationary effect and together with increased AD (AD_0 to AD_1) as a result of increasing incomes, the economy sees long run growth and also price stability at *point B*.

3. Growth and benign deflation: A final possibility is that households hold off on consumption due to the *expectations effect* of falling prices (households hold off on purchases waiting for prices to fall further) and there is negative inflation or *deflation*. In this scenario the increase in factors and productivity shifts SRAS from $SRAS_0$ to $SRAS_1$ and general equilibrium is established at *point C*. Falling price levels are usually associated with falling AD and falling income (such as Japan has seen for long periods) but in this case there is also growth; hence 'benign' (roughly 'kind' or 'benevolent') inflation.

---

1 See Economic News Release at http://www.bls.gov/news.release/prod2.nr0.htm

The shift in LRAS creates an initial negative output gap which over time is closed when…
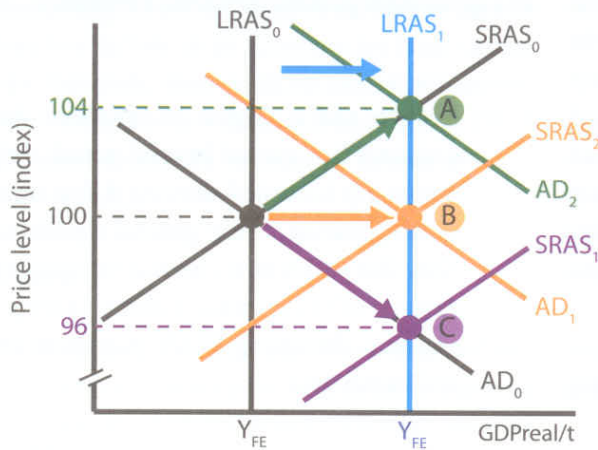
A: …the economy expands in terms of population and increasing economic activity, AD increases from AD_0 to AD_2 and closes what in effect is a negative output gap.

B: …increased productivity shifts SRAS from $SRAS_0$ to $SRAS_2$ while the increase in income also drives up AD from AD_0 to AD_1.

C: …increased/improved factors shift SRAS from $SRAS_0$ to $SRAS_1$ resulting in 'benign' deflation. AD does not increase as households hold of on consumption due to deflationary expectations.

Figure 49.2 – shifts in LRAS and possible scenarios

Land, labour and capital again

Figure 49.3 is pretty much a reprise of the production possibility frontier issues illustrated in Chapter 2. Bundling together human and capital investment on the Y-axis, an economy can increase potential output in the long run be incurring the opportunity cost (sacrifice) of present consumption. This is shown by the movement from **A** to **B**, and the shift of the production possibility frontier from PPF_{SR} to PPF_{LR}. Point **C** is – possibly – a new level of total output.

- more investment in fixed capital such as machines and factories

- road-building, dams, telecommunications and other infrastructural investment

All of the above will shift the PPF outwards in the long run, enabling potentially higher output at point C in the future.
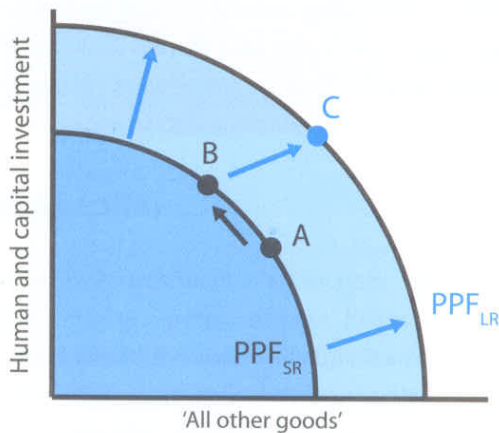
Figure 49.3 – shifts in LRAS and possible scenarios

The movement from A to B can result from:

- improvements in quality and use of land resources

- increased spending in education

A WALK ON THE WEIRD SIDE; DAYLIGHT SAVINGS TIME AND GROWTH

Can daylight savings time actually increase growth rates?! Well, yes, apparently so.

JAPAN'S economy has been in a funk for most of the past two decades. The country has tried fiscal stimulus. It pioneered quantitative easing. Yet still the outlook remains dark. Might giving the public an extra hour of sunlight help?

The 27 countries in the European Union will set their clocks back for winter by one hour on October 31st; Americans do so on November 7th. Japan stands virtually alone among rich countries in not using daylight-saving time (DST), or summer time. The country did so between 1948 and 1951 during the American occupation, but stopped after General MacArthur left Tokyo.

John Alkire of Morgan Stanley believes that adopting DST would mean a new dawn for the Japanese economy. One extra hour of sunlight every evening for seven months would boost domestic consumption, as people leave work for bars, restaurants, shopping and golf. Summer time is credited with reducing traffic accidents and crime; boosting energy efficiency as people use less lighting and heating; and even improving health as people are radiated with vitamin D. "The best part is that it doesn't cost anything," chirps Mr Alkire. "It's a real fiscal stimulus without any money."

Bright idea, Economist, Oct 30th 2010

Consequences of growth

"Once you start thinking about growth it's hard to think about anything else."
Robert Lucas, Nobel laureate

I now raise the bias/subjective flag as I start off with a rather personal note here. I find something disconcerting about the caption 'Consequences of growth'. It is as if one is putting on a new shoe but waiting for the other to drop into the sewer – e.g. one should rush through the positive issues of growth such as poverty reduction, increase in standard of living, increase in choices for people, improvements in quality of live…in order to get to the 'real life' issues of *negative* effects such as environmental sustainability and income distribution. I have found that this is a rather common stance and one with a very heavy implicit *moral* take – it is somehow implied that the 'race for riches' necessitates giving up more 'human' aspects of development found in the 'softer' variables associated with quality of life. Having ploughed through quite a lot of data in the past decade, it becomes ever clearer that as soon as one starts looking at *measurable* and *quantifiable* data on growth and quality of life … well, Lucas (quote above) might have a point![2]

Benefits of growth

Economic growth benefits society and society is made up of people. Have a quick look at circular flow model in Chapter 36 and the connection to the five points below should be clear.

1. **Higher incomes**; households will be able to increase consumption of goods and services. This in itself is an increase in *living standards* as households can enjoy better food, housing and recreation. As shall be seen in Section 4, this has enormous implications for developing countries in terms of poverty reduction.

2. **Lower unemployment**; increases in consumption and thereby aggregate demand *create jobs* in industries.

3. **Government tax revenue**; increases output and expenditure mean that vital tax bases such as expenditure taxes and labour taxes increase *government tax receipts*.

4. **Increased public and merit goods**; increases tax receipts enable governments to provide more *public services* such as infrastructure, education and health care.

5. **Improvements in environment**; while in no way conclusive across a full range of environmental indicators, there is evidence that higher incomes not only *enable* governments to preserve natural environments and take care of precious natural resources, it *incentivises* governments. When a sufficiently high income is achieved people can afford to value the environment more and the increased wealth enables governments to enact environmental legislation and institutions which safeguard the standards.

The correlation (again, not causality!) between national income and standard of living is quite clear and illustrated in Figure 49.4. Using the Human Development Index (HDI, which is comprised of GDP per capita, life expectancy and education) as a metric (= statistic for measuring) of standard of living, it is evident that higher income is correlated to increased living standards. The higher the HDI value, the higher the standard of living – the curve shows that living standards increase dramatically up to a GDP per capita of about USD10,000 after which there is a slower rate of increase.

MACROECONOMICS

[2] Once again, the title of this book is *The Good, the Bad and the Economist* – not *The Good, the Bad and the Ugly*.
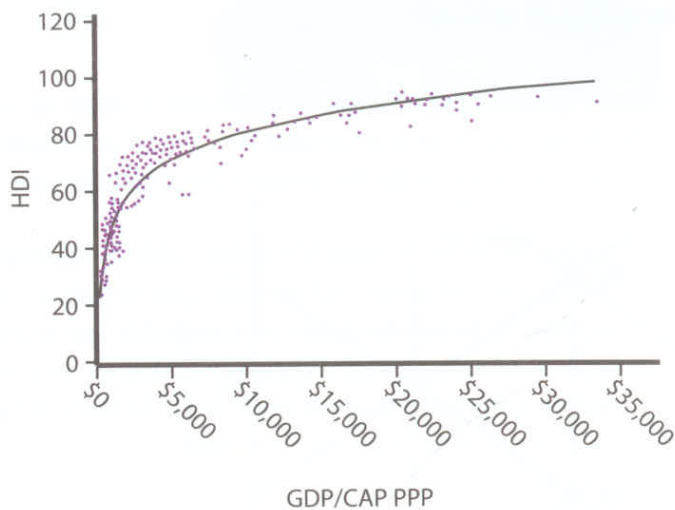
Figure 49.4 GDP per capital and HDI
(Sources; HDI 2008 at UNDP, GDP figures at World Bank)

Possible negative consequences of growth

However repetitive it will sound, everything has an opportunity cost. Growth comes with some strings attached and often in the form of trade-offs.

Inflation

In Chapter 46 we dealt with one of the key trade-offs in macroeconomics, namely the possibility that demand-driven growth can be inflationary. This cornerstone of new-classical theory will be discussed in some detail in Chapters 53 and – for HL – 54. The basic tenet (= rule) is that any increase in aggregate demand and thus growth beyond the long run potential of an economy will result in inflation.

Externalities

Recall that any consumption and/or production which levies costs or benefits on a third party is an externality. Growth will come with such a price tag attached: water and air pollution; traffic noise and congestion; encroachment of wildlife areas; and thousands of other negative results which are the unintended consequences of growth. Economic theory has increasingly sought to include these consequences in tallying up the costs and benefits of economic growth and development. The World Bank's World Development Report (WDR) in 2003 states that a 3% growth rate in the world will quadruple global GDP by 2050.

An example of how income growth affects the demand for goods which have high income elasticities is the growth of car ownership in Mexico. Income elasticity of cars is approximately

2 and the average yearly increase in income in Mexico was 3.4% between 1973 and 1998. During the '90s, the amount of private vehicles had increased by 30% leading to reccurring traffic gridlock and pollution[3].

Population growth and the environment

Even with constant population growth rates, population increases will put additional burdens on the environment. Most estimates of population growth indicate that around 2 billion people will be added to the current global population of 6 billion within the next 50 years – almost all of it in developing countries[4] (see Figure 49.5). This will put an immense strain on our resources and the natural environment, adding to an already alarming list of externalities:5

- **Air:** Greenhouse gases will continue to grow unabated if there is no large scale movement away from the burning of fossil fuels. Seven economies account for 70% of global CO_2 emissions – the US alone, with 4% of the world's population, accounts for 25% of the CO_2 emissions.[6]

- **Water:** There is a real risk that fresh water becomes increasingly scarce in a world where one third of the population already experiences chronic shortages of drinkable water.

- **Land:** Soil degradation – such as erosion and increased salt content – has already affected close to 2 million hectares of land since 1950, 320,000 hectares virtually irreversibly.

- **Forests:** One fifth of tropical forests have been cleared since 1950, which often results in increased desertification (= spread of deserts).

- **Fish stocks:** Over 70% of the world's fish stocks are exploited to the point of carrying capacity – or simply over-exploited.

---

3 The Economist, *All jammed up*, Sep 3rd 1998; and *The World Economy*, A millennial perspective, OECD, table A2-e, page 197)

4 WDR 2003, Chapter 1, page 1

5 These are figures from the WDR 2003, Chapter 1, pages 1 – 5, and it should be noted that they are not uncontested. See for example *The sceptical environmentalist* by Björn Lomborg (op cit).

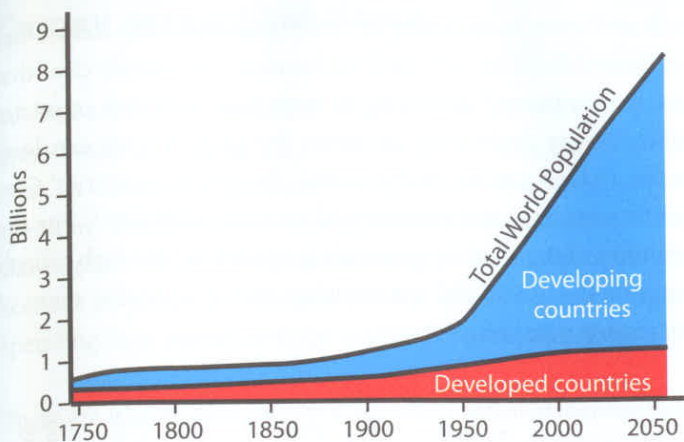6 *Globalization, Growth and Poverty*, World Bank report 2002.

Figure 49.5 World population trend

It isn't difficult to see that growth in the ¾ of the world considered developing will have a considerable impact on the environment if they follow the same path as the developed world. In China, which has averaged some of the highest growth rates of the 1990s, air pollution has reached levels in Beijing and Shanghai which make these two cities the dirtiest in the world. The attraction of jobs and cultural opportunities has led to mass urbanisation and 'mega-cities' such as Mexico City, where population has exceeded 15 million.[7] Urban growth puts additional pressure on infrastructure such as water and sanitation, while inner-city congestion is forcing a number of cities to limit cars.

However ...

Global population is expected to peak at 9 to 10 billion people by 2100 – some 20 to 30% lower than predicted in the 1960s.[8] It is worth noting that are also many *positive externalities* of growth.

- Back roads in distant rural areas built by timber companies give people the ability to commute and get produce to market. (This is clearly evident in the northern Scandinavian forests, where timber roads from the growth period in the 1800s still enable thousands of people to live in the countryside and commute to jobs.)

- An increasing number of firms in high-growth regions will have free medical care, education and housing schemes for employees as a policy of enhancing labour capital.

---

7 During the five years I lived in Mexico City I read of population estimates of 15 million to 25 million. Nobody really knows.

8 *WDR 2003*, Chapter 1, page 4

- Growth in Information Technology centres in India has created a need for – and thus supply of – improved infrastructure such as electricity grids and phone lines which have benefited the region in general.

- Globalisation and trade will spread new technologies – many of which will be far more ecologically sound than older production methods.

There is also evidence that growth – ultimately – has *positive* effects on the environment. While the evidence is very unclear (since there are so many other variables to take into consideration other than income) as to whether growth in poor countries can be tied to environmental improvement, it is quite clear that wealthy countries have seen significant improvement in a number of areas.[9] When people get richer, they *demand* for and can *pay* for a cleaner environment. As incomes increase there are resources available to clean the environment – just witness the improvements over the past 30 or so years; London's air is far better and the Baltic Sea far cleaner than one generation ago. Technology enables efficiency – more Widgets at less pollution – and wealth means that people start to demand other things, like a view unobstructed by garbage; enjoying clean air and fishable lakes… etc. Wealth, technology and disutility then feed off each other. When people get richer they have the 'luxury' of wanting a liveable environment and caring about environmental issues. Technology enables something to be done about it. Wealth allows them to pay for it.[10]

Furthermore, while evidence is not conclusive on environmental performance related to growth in developing countries *specifically*, the OECD has shown broad positive correlation between growth and income in countries *generally*. Figure 49.6 shows that positive correlation between income and environmental performance (a composite index of a number of environmental measures) can be seen for a range of countries.

---

9 See for example *Globalization, growth and poverty*, World Bank 2002, page 130 ff

10 I just needed to look out my window back home in Sweden, at Öresund, the strait between Sweden and Denmark, for a good example. Just 25 years ago, raw sewage went straight into the sea. The strait was so polluted that swimming bans were commonplace, fish stocks were severely depleted, and the bottom of the sea was predicted to be unable to sustain life by the year 2000. By late May of every year, my IB2s were down post-exam frolicking on the beach and the heartier specimens were jumping in the water (14°C) – while sports fishermen cast for wild salmon and sea-trout. The wealthy citizens of Sweden and Denmark simply cleaned up their act – and the sea – by installing waste removal plants and implementing strict legislation on pollution. They were able to do this as they were *wealthy* and could afford the cleaning technology.

MACROECONOMICS

Other evidence – put forward by the World Bank – suggests that growth which is primarily trade-based has not created 'pollution havens' due to multi-national companies seeking out countries with low environmental standards.[11]
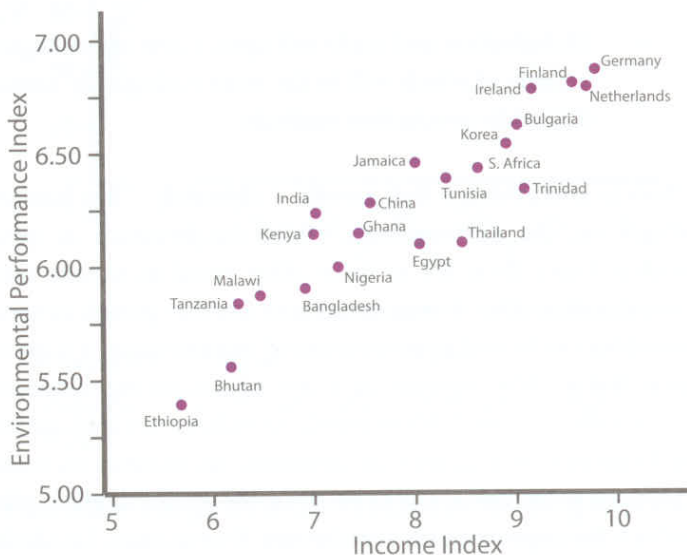

Figure 49.6 Income and environmental performance

Source: *Open markets matter*, OECD policy brief October 1999, page 9

Poverty is, arguably, a worse threat to the environment than growth.[12] Poor people living on ever lower-yielding land will be forced to use marginal resources for survival. They know full well that the depletion and destruction of natural resources worsen future living standards but simply do not have a choice. Poverty also leads to population growth as children are needed for labour to work on land suffering from diminishing returns. Children also provide a 'social security net' in a society where there is scant availability of pensions, health benefits and social security benefits. The forces of increasing numbers of people living off depleting resources create yet another vicious cycle: poverty → population growth → lower productivity and land yields per capita → poverty. Lifting people out of poverty is increasingly viewed as the most effective way of improving the environment.

Income distribution

"A society can be *Pareto optimal* and still be perfectly disgusting." Amartya Sen

There has often been a heated debate as to whether global income distribution is worsening, i.e. whether the 'gap' between

rich and poor is widening in many countries. The debate has quietened down recently as it has become increasingly clear that the 'gap' is indeed increasing in both poor and rich countries alike. Saving developing countries for Section 5 we can look at the richest nations in the world, the OECD countries' Gini coefficients as a measurement of income inequality.[13] For the majority of the wealthy countries in the world, the high growth rates of the 1990s did not translate into a equitable share of increasing prosperity.
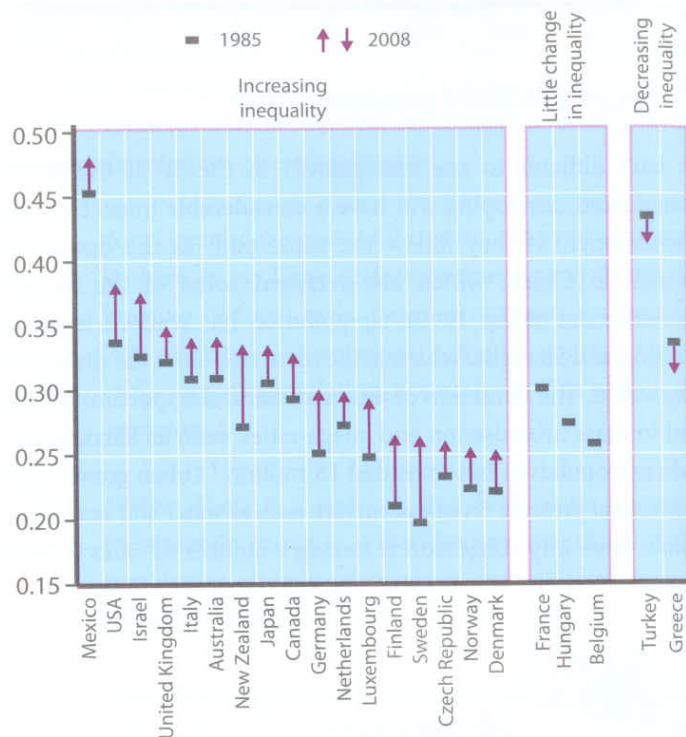

Figure 49.7 Income inequality in OECD countries 1985 to 2008

However ...

Widening income gaps do NOT mean 'the rich get richer and the poor get poorer'. Yes, this is possible but by no means certain. Case in point, look up the US and Sweden over the period from 1985 to 2005. Exactly; first one needs to adjust for inflation and then one needs to see whether the increased disparity is because the lowest income groups have increased their incomes at a slower rate than higher income groups.

---

11 *Economic man, cleaner planet*, The Economist, September 27th 2001

12 See for example Moore, page 89; and Legrain page 244

13 OECD is the Organisation of Economic Cooperation and Development. The 34 members are mostly high income countries. The Gini coefficient can take any value between zero and 100. It measures the 'spread' of a nation's income – the higher the value the more unequal is the distribution of income.

MACROECONOMICS

Current account imbalance

We will look at the current account in balance of payments in great detail in Chapter 70 so for the time being just think of it as an accounting method for export revenues and import spending. Traditionally a 'strong' current account is one which shows greater export revenue than import spending – a current account surplus. An 'imbalance' would thus be greater import spending than export revenue – a current account deficit.

There are two mechanisms at work when national income rises in an economy which both serve to worsen the current account; the *relative inflation effect* and the *income effect*.

- Rising income in the Home Economy which is demand-driven primarily – i.e. the result of increased aggregate demand – is inflationary. *Ceteris paribus* (no change in exchange rates or the inflation rates of trade partners) this means that there is an increase in the inflation rate relative to trade partners. The increase in relative inflation makes exports more expensive for trade partners and imports relatively cheaper for households in the Home Economy. Falling export revenue and increased import spending can therefore worsen the current account.

- In conjunction with the above, rising incomes in the Home Economy lead households to increase their overall consumption. Some of these goods and services are produced in trade partners' economies. This increases import spending and can worsen the current account.

HL extension

Calculating growth from data

Figures 49.9 and 49.10 below give data on GDP for the UK over a six year period and for China over the same period. Note that the latest year (2010) is the base year.

United Kingdom	
GDP, current prices, billion $US	GDP deflator (base year 2010)
2005 2282.9	87
2006 2447.7	90
2007 2812.0	92
2008 2679.0	95
2009 2182.4	97
2010 2250.2	100

Figure 49.8 – GDP in current USDs and GDP deflator, UK, 2005 – 2010

Source: World Economic Outlook, September 2011 and HM Treasury at http://www.hm-treasury.gov.uk/data_gdp_fig.htm

1. Calculate the rate of growth for each five year period (e.g. 2005 to 2006, 2006 to 2007 … etc).

2. Calculate real GDP for each year.

3. Calculate real growth rates over the period.

	China	
	GDP, current prices, billion $US	GDP, current PPP dollars, bln
2005	2256.9	5364.3
2006	2712.9	6240.8
2007	3494.2	7333.8
2008	4520.0	8216.0
2009	4990.5	9068.2
2010	5878.3	10119.9

Figure 49.10– GDP in current and PPP adjusted USDs, China, 2005 – 2010

Source: World Economic Outlook, September 2011

1. When you calculate the (nominal) growth rate, it is evident that China's rate of growth is considerably higher than the US. There are two 'howevers'. Which?

2. The column to the right shows GDP in purchasing power parity (PPP) dollars. This means that GDP values have been adjusted for what a US dollar can actually buy. Why are the PPP values for GDP higher for China? (See Chapter 81.)

Summary & revision

1. **Growth** is defined as an increased in real GDP over a period of time – usually one year.

2. Growth in the long run occurs due to increased quantity and quality of factors:

   a. **Natural factors** such as minerals, oil and arable land.

   b. **Physical capital** due to investment, such as machines and infrastructure.

   c. **Human capital** arising from improvements in education and increased skill base of the labour force.

3. The **productivity of labour** increases due to new technology, increased division of labour, better telecommunications, research and development and the application of motivational techniques in firms.

4. **Positive aspects of growth** include improved living standards, lower unemployment, increased government tax receipts, provision of more public services and possible improvements in environment.

5. **Negative aspects of growth** include rising inflation rates if the increase in income is primarily demand-driven, externalities such as pollution, sustainability problems when natural resources are depleted/destroyed, widening gaps between richer and poorer households, and worsening of the current account.

MACROECONOMICS

50. Low Unemployment

Key concepts:

- Definition of unemployment
- Calculation of unemployment
- Difficulties in measuring unemployment
- Consequences of unemployment
 - Social costs
  - Economic costs

HL extension:

- Calculating unemployment from data

Newlan's Truism: An 'acceptable' level of unemployment means that the government economist to whom it is acceptable still has a job.

Definition of unemployment

The concept of **full employment** has so far been used in conjunction with the long run aggregate supply curve, where long run potential output is also the full employment level of output. Full employment does not mean that there is 'zero **unemployment**' but rather that all of the people willing and able to work have jobs at the going wage rate. Putting things simply – initially – full employment is the quantity of the total labour force employed when the labour market is in equilibrium. Figure 50.1 illustrates this, where the demand for labour and supply of labour at the real wage level W_0 creates full employment at FE.[1]
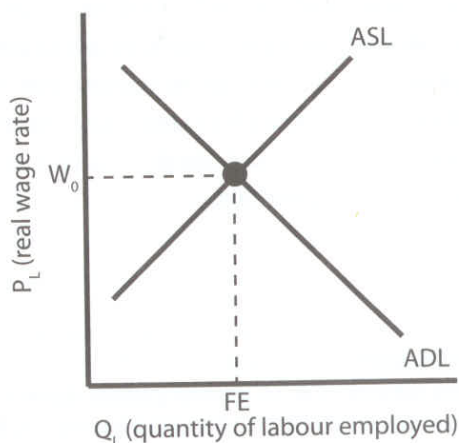
The aggregate supply of labour (ASL) in an economy will be upward sloping, showing that labourers will have an incentive to increase their working hours if real wages rise and also that employers will seek to induce workers to work overtime by paying more. The aggregate demand curve for labour (ADL) is downward sloping, showing that firms will have an incentive to use more labour – substituting capital – as the real wage rate falls. At a real wage rate of W_0 there is full employment at FE.

Figure 50.1 Equilibrium labour market

The concept of full employment deals with the sum total of all labour markets, the *aggregate supply/demand for labour*, ASL and ADL in Figure 50.1. Since employment (and thus unemployment) is a stock concept, at any given point in time there will be people between jobs and entering/leaving the labour market. It is likely that a number of individual labour markets, say dentists, have an excess of demand while there is an excess supply for computer programmers. Full employment therefore does not mean 'everyone has a job', but that in total the supply of labour equals demand. This is why full employment is defined as equilibrium on the *total* labour market rather than individual labour markets.

---

1 Again, we deal solely with **real** values. Real wages are average wage levels in the economy adjusted for inflation, i.e. set in base year values.

Underemployment

An initial weakness of the concept of full employment is the simple fact that any number of people who are technically employed will in fact be working part time or in odd jobs and therefore in effect are suffering from **underemployment**.

- **Women**, for example, are often caught in an underemployment trap due to traditional roles of child care; while many might be looking for close to full time employment. As the children grow up women often have difficulty finding increased work hours.

- People working **part-time**, such as handymen and odd-jobbers will frequently be underemployed over the course of the year, since such labour demand is highly varied.

- Finally, **developing countries** often have high levels of underemployment since rural households and families will soak up a good deal of what otherwise would be counted as unemployment. This has been accentuated by high levels of population growth which add to the labour force and lowers the productivity of labour.

Definitions: 'Unemployment, full employment and underemployment'

Unemployment is the number of people in the labour force – e.g. able and willing to work – not holding a job. It is usually expressed as a percentage, for example 400,000 unemployed out of a total labour force of 5,000,000 is 8% unemployment.

Full employment is when there is equilibrium on the aggregate labour market, i.e. the total number of hours demanded by firms corresponds to the total number of hours supplied by households at the going real wage rate.

Underemployment is when people who are offering their services on the labour market can not find full time work.

Calculation of unemployment

"A recession is when your neighbour is out of work. A depression is when you are out of work".
Harry Truman

Defining and then measuring the unemployment rate looks straightforward and simple; take the number of unemployed people at a given point in time in an economy and divide the amount by the work force. This seemingly simple stock concept comes with quite a few 'howevers', but let's start off by way of a basic example. Figure 50.2 shows the unemployment figures in Canada in December 2008.[2]
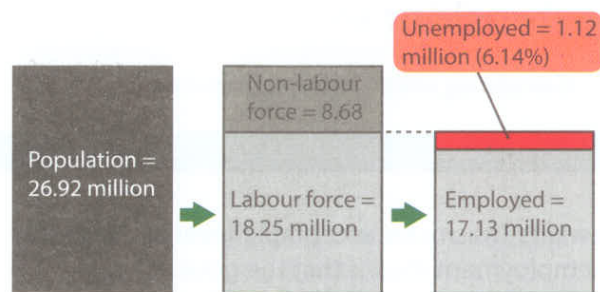

Figure 50.2 Unemployment in Canada 2008

Out of a total of 26.92 million citizens, the Canadian labour force is just over 18 million workers, or 67.8% of the population – also known as the **participation rate**.[3]

So who are the other 8.68 million Canadians in Figure 50.2 who are not either felling pine trees (for lumber) or sticking taps into maple trees (to make delicious maple syrup)? Well, the labour force is defined as those **above the age of 16 who are considered to be economically active**, i.e. are either *seeking* employment or *have* employment. This excludes a number of people other than under-16s such as military personnel, retirees, parents on child leave, physically and mentally disabled … and perhaps students, those who do not want to work, people in government re-training programmes – and anyone unfamiliar with Paul Bunyan[4]. It depends on how both the numerator (unemployed) and denominator (labour force) are defined and measured – and this differs between countries.

---

2   Official government figures at http://www.statcan.ca/english/Pgdb/labor20a.htm

3   I note with interest that in updating these figures from the previous edition where figures for 2003 were used, that while the population has increased from 25.25 to 26.9 million and the labour force 17.2 to 18.25 million, the participation rate is identical at 67.8%.

4   Joking! Paul Bunyan was a mythical North American lumberjack of bedtime story type. Recommended reading.

MACROECONOMICS

Difficulties in measuring unemployment

The message here is; be careful in comparing official unemployment rates for different countries! Each country will have their own methodology for defining the labour force – and also those considered to be unemployed. The problems arising in trying to correctly estimate the rate of unemployment are of two main types; *definitional* problems and *accounting* problems.

- **Definitional problems** arise for example in defining whether those with disabilities and long term illnesses, unemployed workers close to retirement age and those who no longer bother to seek work (known as **discouraged workers**) should be counted as part of the labour force or not. In many developing nations there is also the problem of weak censuses whereby people are not even part of the population count.

- **Accounting problems** arise when unemployment figures are primarily based on those who have registered as seeking employment at employment agencies and those receiving unemployment benefits (claimant count); there could be a large number of people who in reality are unemployed but are 'hidden' in job training schemes, youth employment/training programs and people who are not even part of the formal economy.

One way to deal with the problems of definition of and accounting for unemployment is by standardising the measurements and then doing a **survey** of the population. This is what the International Labour Organisation (**ILO**) does. The results are commonly used for international comparisons. The ILO defines unemployment as people who are without a job (= work less than one hour a week), have actively sought employment in the past four weeks and are able to start a new job within two weeks.[5] A cross section representing the entire country is then surveyed via forms and interviews, whereupon the results are statistically organised to show the total unemployment rate. The Organisation of Economic Cooperation and Development (OECD) uses the ILO standard, giving the unemployment rates shown in Figure 50.3.

---

5 My experience in Mexico re the ILO definition; it ain't working! One of my students, Ximena, did a brilliant extended essay on actual unemployment in Acapulco, a tourist resort city on the west coast. Since so many labourers are employed in the tourism industry and often rely on occasional odd jobs, the official rate of unemployment is around 2%. In reality the figure is closer to 25%. The results were stark illustration of the weaknesses in unemployment figures.

Figure 50.3 Unemployment in the OECD 2010

Consequences of unemployment

Having high levels of employment is perhaps the foremost macro objective since unemployment is strongly associated with societal costs. While many of these costs are naturally primarily felt by the unemployed themselves – primarily loss of income and living standards – a good many costs are inflicted on society. The costs of unemployment are often put in terms of economic and social costs – which of course overlap each other.

Social costs

The above highlights the allocative waste involved in unemployment; labour and labour capital is a most valuable resource in improving living standards in an economy. These economic costs quite naturally glide into social costs. I often tell my people to ask their parents how they 'identify' themselves, i.e. how they see themselves 'from the outside'. Is it 'Jan the parent'; 'Jan the doctor'; 'Jan the fly-fisherman'; etc. Odds are in favour of people identifying themselves in terms of *occupation*, here; 'Jan the doctor'. Modern societies (frequently in the form of parents) often put heavy emphasis on 'getting a job and being

able to support yourself' and career-building. I recall a story about an upper-level manager in Japan who was laid-off during a recession and couldn't face letting his family, friends and neighbours find out – so he continued to get up at 6:00 in the morning to 'go to work', i.e. sit in the park and read the papers.[6] In many societies there is enormous social stigma (= dishonour) attached to being unemployed.[7] I venture to say that Figure 50.4 needs no further comment.
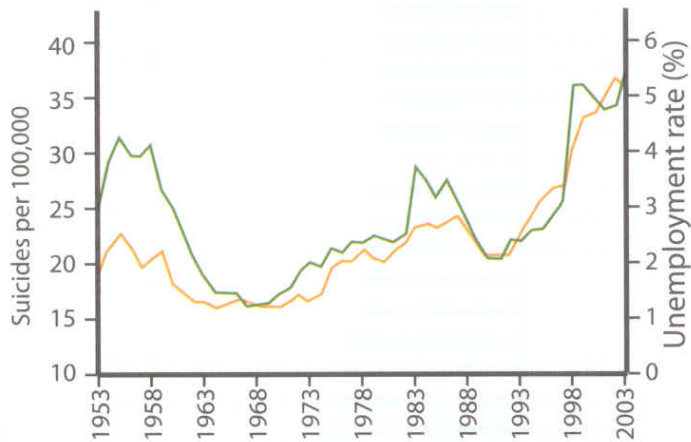

Figure 50.4 Correlation between unemployment and suicide – Japan

(Source: Japanese Ministry of Health, Labor and Welfare – see www.mhlw.go.jp/english)

The social effects are perhaps even more damaging than the economic effects – and this is not touchy-feely-economics, but a hard-line view of economic realities. Just ask yourself who will be the most productive/constructive member of the community; a proud, committed, confidant and dignified person – or someone with low self-image and little sense of self-worth. Yes, it is actually that simple; being unemployed is highly value-laden and will damage those who are subjected to it for longer time periods and by extension this will damage society. Macroeconomic studies consistently show correlation between unemployment rates and domestic violence, crime, drug and alcohol abuse, broken families and mistreated children, and general mental and 'spiritual' (for lack of a better word) ill health leading to higher suicide rates and depression – all of which are quite evidently economically sub-optimal.[8]

---

6   A couple of my Japanese students, Buddha and Shota, read this and commented; "Well, yes, of course. Eh?" Oh well, at least they bring me sake.

7   I recommend the film *Falling down* by *Robert Altman* for a most illustrative case in point.

8   I was invited out to dinner by parents quite often when I was living in Mexico. I recall one evening where the father and I sat after dinner puffing cigars, drinking Cognac and discussing the on-going plight of the world economy (January 2008). This man was a rather powerful individual in the business

Economic costs

There are numerous economic costs involved in having people unemployed:

- Perhaps the most intuitive cost of unemployment is the **loss of output** in the economy resulting from idle factors of production. While these costs might be relatively lenient in the short run, in the long run there will be a number of negative aspects which arise.

- Lower income in the economy will diminish the **tax base** and decrease income tax receipts for government.

- This will put an additional burden on the budget which would in all probability already be under pressure due to **increased transfer payments** in the form of unemployment and social benefits.

- The use of government funds in this manner might mean **opportunity costs** in the form of less money available for schools, health care and roads. Or even if the government keeps planned government spending constant it might borrow money to fund deficit spending which could **crowd out** private investment (see Chapter 57). Government might instead increase taxes in order to make up for the loss of tax receipts, in which case there will be a burden on the tax payer.

- Longer periods of high employment also make it **difficult for labour market entrants** (e.g. young people looking for initial employment) to get into the market, since employers will have the 'pick of the crop' and naturally look to the experienced workers first. Finally, long term unemployment will inevitably erode the skills and abilities of experienced workers (deskilling) and reduce the economy's ability to get back on its feet later on.

The last point was driven home rather scarily during the economic crises starting in 2007. Business Week took a good hard look at previous recessions and compared them to the then on-going crisis in 2010. The result is quite alarming as shown in Figure 50.4; for each consecutive downturn since the 1980s, it has taken increasingly longer for the US economy to

world and heavily in the know. I mean, several times during the course of conversation there were snippets like "...and I told [Mexican President] Calderon that ..." He, as I, spoke anxiously about the increase in very nasty criminality that will doubtlessly ensue when the full force of falling national income and rising unemployment hits Mexico. He has since in fact left the country for fear of his family's safety.

'bounce back' to the initial rate of unemployment at the start of recession. There is increasing evidence that even during a recovery unemployment rates will linger far above the long run historical averages of 5.5 to 6.5%.
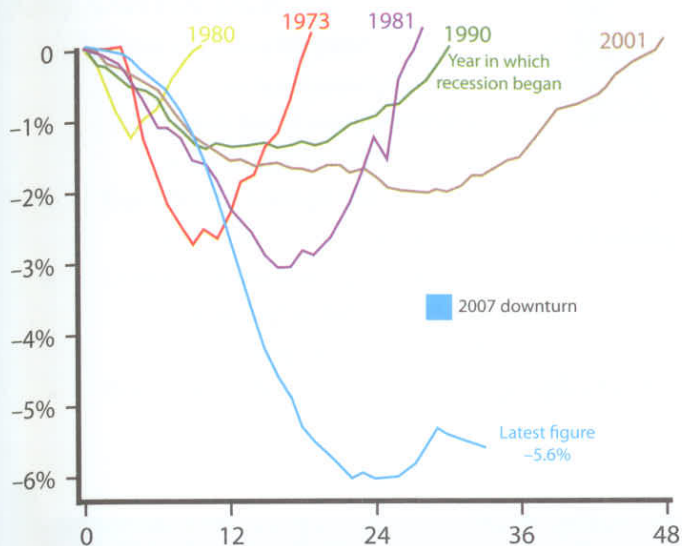

Figure 50.4 Unemployment changes over past six recessions in the US

(Source: BW, Oct 18 – 24 2010, "A $14.5 trillion economy stuck in neutral")

HL extensions

Revise the core definition of unemployment and simply keep your eye on which data you need to arrive at unemployment: 1) number of people out of work; 2) size of the total labour force.

Calculating unemployment from data

Let's not make it *too* easy. Below I've built a Home Economy with a some data on population. Your mission is to arrive at what the unemployment rate is. Notice: I've built in some data that you don't need![9]
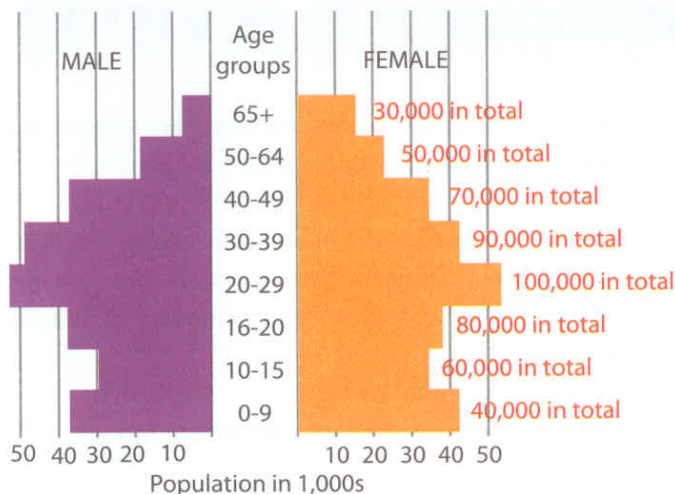
Data set I: Population pyramid

Data set II: Population break-down

- A survey indicates that 24,000 people are currently unemployed

- There are 50,000 kids in high school

- 20,000 are unavailable for work for health reasons

- 10,000 are doing military service

- An additional 10,000 people are at university

- 20,000 people are taking care of the home but are registered at the employment agency

- 30,000 people are working 50%

- The retirement age is 65

---

[9] That's one of the main reason's I became a teacher, to irritate people.

Summary & revision

1. **Unemployment** is defined as the total percentage of the total labour force not holding a job.

2. The **total labour force** does not include under-16s, over-65s, disabled or otherwise unable to work, people doing military service and students in secondary or tertiary education.

3. **Difficulties in measuring unemployment** include

    a. **Definitional problems** of estimating the labour force; which persons to include as long term ill or disabled, discouraged workers not bothering to seek employment and people who are simply not registered in a census

    b. **Accounting problems**; claimant counts (e.g. registered at the unemployment office) are seldom accurate in measuring those not holding jobs, hidden unemployment due to government workshops and re-training schemes and large informal sectors.

4. The costs of unemployment can be divided into social and economic costs:

    a. **Social costs** include increased crime rates, suicide, drug and alcohol abuse and depression.

    b. **Economic costs** include loss of output and income, lower tax revenues to government, increased burden on governments in the form of unemployment and social benefits to jobless and deskilling of labour making it harder to rejoin the labour market.

51. Types and Causes of Unemployment

Key concepts:

- Types of unemployment
    - o Structural unemployment
    - o Frictional unemployment
    - o Seasonal unemployment
- Equilibrium unemployment
    - o Full employment and natural rate of unemployment
- Disequilibrium unemployment
    - o Cyclical unemployment
    - o Real wage unemployment
- Government policies to deal with unemployment

Since unemployment is a stock concept, the level will change continuously as people enter and leave the labour force. An increase or decrease in population, school-leavers, and people simply returning to the job market after illness or paternity/maternity leave will serve to increase/decrease the size of the labour pool. Conversely, assuming a constant labour force, there will be people from the labour force gaining and losing jobs for any number of reasons. So if more people gain jobs than lose jobs unemployment will fall.

Types of unemployment

This ongoing process of in- and out-flows into the labour market and available jobs will necessarily mean a degree of unemployment will always exist at any given point in time; this is *equilibrium unemployment* and consists of three main types; **structural**, **frictional** and **seasonal** unemployment.

Structural unemployment

The most difficult type of unemployment to deal with is when it is 'built-in' to society by forces inherent to the economy itself. **Structural unemployment** is the mismatch of available labour skills and the demands of the economy. The main forces affecting unemployment levels here are *labour mobility*, *declining* ('sunset') *industries* and job redundancies resulting from *technological advances* in production. Any changes of these variables will give rise to structural unemployment since labour is configured to match certain demands in the labour

market. There will be job losses when labour demands change, often resulting in very painful long run unemployment for large numbers of workers *in a specific industry* – e.g. there is structural change in the economy whereby demand falls (permanently) for labour in a certain industry.
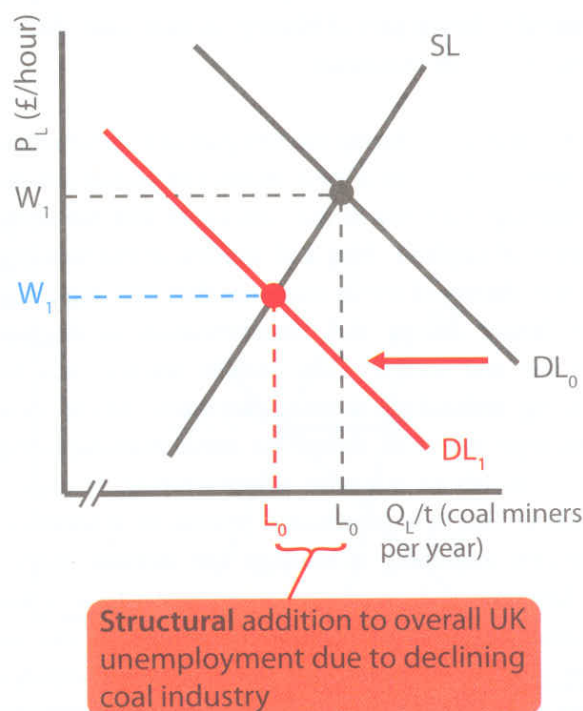

Figure 51.1 Structural unemployment amongst coal miners, Northumberland UK, 1980s

Figure 51.1 shows the resulting increase in unemployment as demand for coal miners fell in the UK during the 1980s. The deep mine pits in the Northumberland region could not compete with far cheaper surface coal in China and Indonesia and as production shifted away from the UK and the newly elected conservative government of Margaret Thatcher stopped subsidising production, demand for coal miners plummeted (DL_0 to DL_1). In 1984 coal miners comprised about 1% of the total labour force and 20 years later less than 0.02%.[1] Structural unemployment is one of the most serious macroeconomic problems facing nations today, and three (often overlapping) types of structural unemployment can be identified.

1. **Regional unemployment** happens when a main industry in a region suffers from permanent loss of jobs, such as the fishing industry along the Norwegian coast (caused by depleted fish stocks); the steel industry in North America's so-called 'Rust Belt' (caused by increased global competitiveness); and coal mining in the Northumberland region of England.

2. Very similar, and indeed often arising simultaneously, is **sectoral unemployment**. This is when a specific sector of the economy suffers from lengthy and often permanent decline. Industries producing horse-drawn buggies, mechanical adding machines and typewriters, and slide rules were all once commonplace but are now largely relegated to being part of history. In each case, skilled labour was made redundant.

3. In 1811 and 1812, during the first part of the Industrial Revolution, there was a great number of violent attacks on weaving machines in the Lancashire and Yorkshire districts of England. Men and women in the weaving industry destroyed textile machines that were replacing their labour during this transformation of English economy and society.[2] This violent reaction was to on-going **technological unemployment**, where new technology in textiles shifted the entire manufacturing base from labour intensive manufacturing to capital intensive. This is a continuous process in economies, since ever-improving technology will increase output per labourer and in many cases create large scale redundancies amongst professional groups. As one of many possible examples, just think of how computer

1   See a very good article at http://www.geographyinthenews.rgs. org/news/article/?id=273
2   These became known as 'Luddites' in reference to a fictitious 'General Ludd' that the workers professed to follow. Today the term Luddite is used rather derogatorily as someone who is 'techno-phobic' and 'backward-striving'.

technology has vastly increased the capacity of putting together a newspaper; thousands of highly trained professional typesetters were made redundant within a few years.

Where have all the draymen gone?

Frictional unemployment

Labourers leaving/losing one job will mostly set out to find another. This is **frictional** (or **search**) **unemployment** and is mostly short term. The speed with which job seekers are able to find new employment depends on their work skills and education, plus the needs of the labour market. In addition to these market forces, there are a number of 'interventionist' variables which will affect the ability/willingness of employers to hire and labourers to accept jobs. For example if labour law makes it difficult for employers to lay off workers then there will be greater care – and time spent – in hiring new workers. If unemployment benefits are high, the unemployed will have relatively low incentives in looking for jobs. (See *Supply side policies* in Chapter 60.) In addition to this, the availability and efficiency of job centres and unemployment agencies will affect time between jobs.

Seasonal unemployment

Waiters in holiday resorts; ski instructors; construction workers on North Sea oil platforms will all have to deal with longer periods of inactivity, giving rise to varying patterns of **seasonal unemployment**. Other than these workers finding backup jobs in off-seasons, there is little to be done about this type of unemployment. This type of unemployment will often overlap regional unemployment.

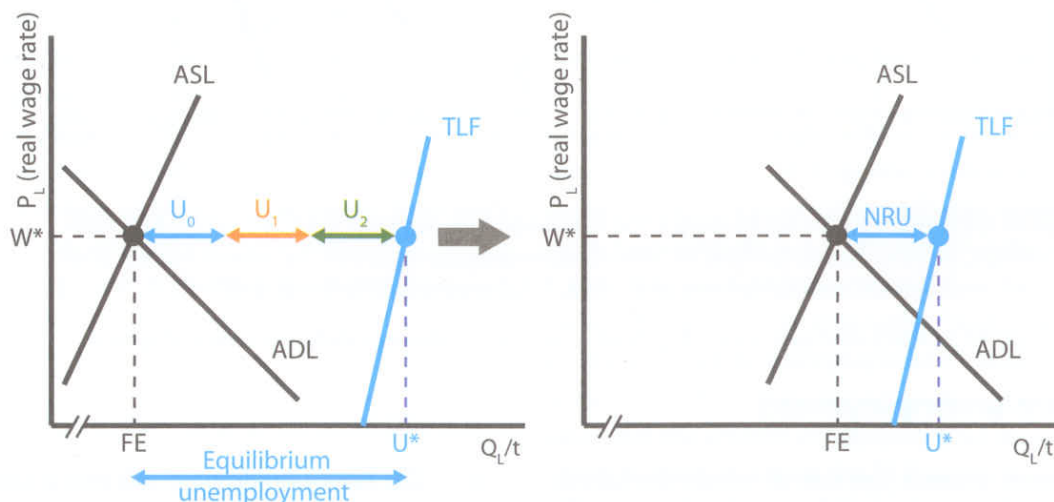## Equilibrium unemployment – why we are different from chickens

Chickens rarely – if ever – volunteer their services on the food market.[3] Labour is a little different from the market for chickens (or any other good, of course) in that the 'good' on offer (e.g. labour) has a mind of its own and will therefore *cause its own supply*. In other words, there will be a difference between the

---

3 But the genetic modifiers are probably working on it. Kamikaze chickens, coming up.

total *potential* supply of labour – the labour force – and the *actual* supply of labour. While the illustration of the market for chickens simply shows the supply and demand at any given price, it is far more interesting when one is illustrating the labour market to show *how many* chic….eh, labourers are NOT part of supply. The difference between the people supplying their labour and the total labour force is of course unemployment.

Equilibrium unemployment – the natural rate of unemployment

Figure 51.2, diagrams *I* and *II* illustrate how *structural, frictional* and *seasonal* unemployment together make up the **natural rate of unemployment**. The aggregate supply of labour, **ASL**, shows the quantity of labour supplied at any given wage level, while the total labour force, **TLF**, shows the potential amount of labour available if everyone who offered their services had a job. (Think of the **ASL** curve as the '*job acceptance curve*' – it shows the willingness of labourers to accept a job at a give real wage rate.) The **TLF** curve is upward sloping since higher wages would induce more people to enter the labour force, for example recent retirees and discouraged workers who would find it increasingly worth their while to offer their labour on the market. Notice that the **TLF** curve is steeper than the **ASL** curve, indicating that as the real wage rate increases, ever fewer people will spend time unemployed so the distance between the two curves decreases.

There will always be a degree of unemployment – even when the labour market has cleared. **Equilibrium unemployment** is therefore the same as the **natural rate of unemployment (NRU)** shown as the difference between FE⇔U*.

Figure 51.2 Equilibrium in the labour market

The curves don't meet since there will always be someone unwilling to accept a job no matter what the wage level.[4] In Figure 51.1a, the *labour market is in equilibrium* at the full employment level, **FE**. At the going wage rate of **W***, there will still be a number of people in the labour force who are unemployed; the *structurally unemployed* (U_0), the *frictionally unemployed* (U_1) and the *seasonally unemployed* (U_2). The sum of unemployment at the market clearing real wage level **W*** is FE $\Leftrightarrow$ **U***. This is the **equilibrium level of unemployment**.

Figure 51.1b (note; different scale) shows the sum of the three unemployment types. Structural, frictional and seasonal unemployment together comprise what rather contradictorily is called the full employment level of unemployment. A proportion, e.g. percentage, of the total labour force which chooses not to accept jobs at the going wage rate will vary but exist at *all wage levels*; thus it is also the **natural rate of unemployment**.

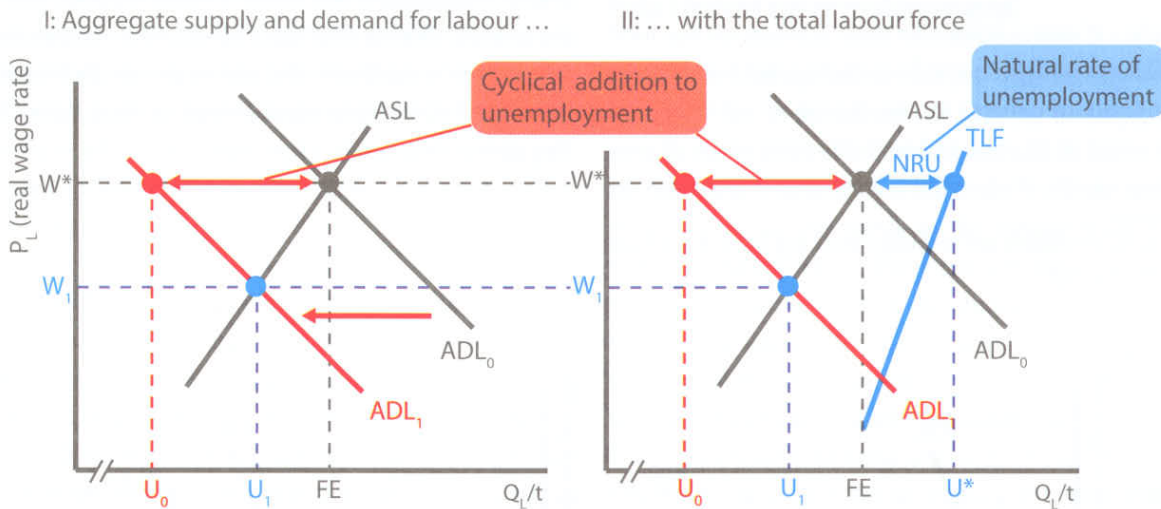4   His name is *Lars Swahn*. I went to university with him.

Definition: 'Full employment and equilibrium/voluntary/natural unemployment'

I know all the terminology gets confusing, so allow me to do the definitions via another syllogism:

1) If everyone in the labour force who wants a job at the going wage rate has a job, there is **full employment**.

2) Since these people accept the real wage rate, ASL equals ADL and the labour market is in equilibrium – any existing unemployment is thus **equilibrium unemployment**.

3) Since there is always an element of voluntary unemployment in the economy, it is natural. The percentage of workers voluntarily unemployed is the **natural rate of unemployment**.

Hence: Full employment = equilibrium unemployment = natural rate of unemployment.

I: Aggregate supply and demand for labour …   II: … with the total labour force

Cyclical addition to unemployment

Natural rate of unemployment

Due to downward stickiness of wages, the real wage rate does not fall to W_1 when **ADL** falls from ADL_0 to ADL_1 but remains at **W***. This creates disequilibrium on the labour market and cyclical unemployment of $U_0 \Leftrightarrow FE$. Total unemployment is $U_0 \Leftrightarrow U^*$.

Figure 51.3 Disequilibrium unemployment – cyclical or demand-deficient unemployment

Disequilibrium unemployment

When unemployment exceeds the rate at which the labour market is in equilibrium, there is *disequilibrium* unemployment. Economic theory identifies two main reasons for disequilibrium unemployment and the debate over which is a more accurate view is frequently rather high-pitched and politicised – this is at the core of the Keynesian – new-classical debate.

- The *Keynesian/demand-side view* is that disequilibrium unemployment is caused by **too little demand** in the economy.

- The *new-classical/supply-side view* is that disequilibrium unemployment is caused by **market imperfections** such as minimum wages and strong unions.

Cyclical unemployment (or 'demand-deficient' unemployment)

We start off by looking at cyclical (or demand-deficient) unemployment. In this view, unemployment is strongly linked to phases in the business cycle, as the demand for labour is derived from the demand for goods and services. This is known as **cyclical unemployment** since it is connected to cyclical variations of economic activity. When total unemployment is higher than the natural rate of unemployment there is a *cyclical addition* to total unemployment caused by relatively low aggregate demand.

Diagram I in Figure 51.3 utilises the aggregate demand and aggregate supply curves for labour to illustrate how a change in the demand for labour during a recessionary period causes *disequilibrium* on the labour market, i.e. a degree of **cyclical unemployment**. The decrease in the demand for labour from ADL_0 to ADL_1 adds $U_0 \Leftrightarrow FE$ unemployment to the previous (natural rate of) unemployment $FE \Leftrightarrow U^*$ shown in diagram II. The question you should now be asking yourself is *why* the real wage rate doesn't fall to W_1 and create a new equilibrium level of unemployment lower than the rate at W^*.

Here's a clue; cyclical/demand deficient unemployment is also known as **Keynesian unemployment**! Recall that Keynesian economics views markets as imperfectly functioning in general, and that labour markets specifically suffer from downward stickiness. The concept of cyclical unemployment, in accordance with Keynesian assumptions, means that real wages will not fall in the short run and the market will be in disequilibrium. In other words, since labourers will be highly unlikely to accept lower wages (and firms will also be reluctant to lower them) the real wage rate will remain at W^* and create an excess supply of labour at the going wage rate. While the labour market might ultimately clear, the proportion of people not accepting jobs at lower rates might last for some time, with unemployment rates above the market clearing level of W_1. Keynesian economics looks upon markets as *inherently unstable/imperfect* – which explains the propensity towards government intervention.

DANGER MINES # WARNING !

Using labour market diagrams

Here are a few common mistakes you will want to avoid – plus a few comments on my labour market diagrams.

Common errors:

- A common error in using labour diagrams is to mis-label the axes. Quite frequently – when exam stress sets in – students will use 'AD' rather than 'ADL'.

- Another common error is to put 'P' on the P-axis rather than the correct 'P_L' or 'Real average wage rate'.

- Include relevant assumptions! Here, you should include whether you are assuming real wages to be downward sticky or not.

Comments on my use of diagrams to illustrate unemployment:

I know of no secondary school textbook which consistently uses the TLF curve in diagrams seeking to illustrate changes in ASL or ADL. Commonly, textbooks briefly explain labour market equilibrium – e.g. the natural rate of unemployment – and then use a rather simplistic diagram omitting the TLF curve. My question is invariably: "Right, now how can you see the quantity of unemployment?!" Well ... no, as a matter of fact. A shift in ADL shows the change in people who *have* jobs – not the amount of labourers who do *not* have jobs. Thus I have decided to retain the TLF curve in all diagrams. **I am telling you the above only to point out that you will NOT be penalised in your exams or IA by using the more common version without the TLF curve.**

Aggregate demand and demand for labour

The relationship between aggregate demand and the demand for labour should be fairly clear cut; a rise in aggregate demand will lead to an increase in the demand for labour and vice versa – the *demand for labour is derived from the demand for goods and services*. Figure 51.4 shows how a decrease in aggregate demand during a recession will affect the aggregate labour market.

Diagram I: The business cycle

II: Goods market

III: Labour market

Figure 51.4 Disequilibrium
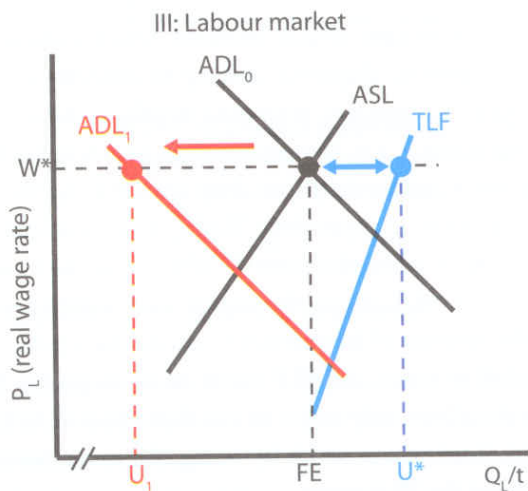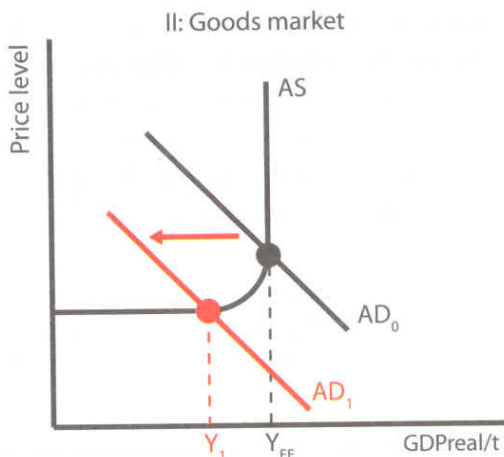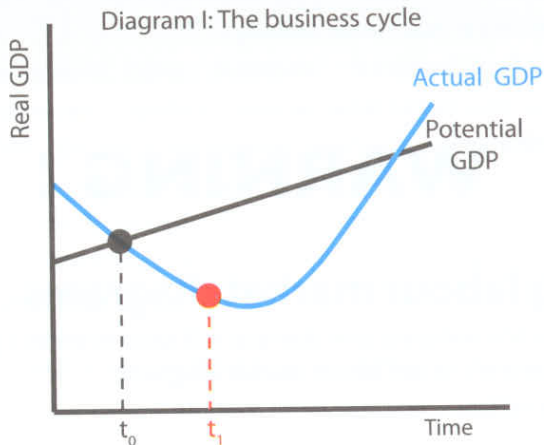
- During a recession (Figure 51.4, red section in diagram I) aggregate demand falls which is shown in diagram II as a decrease in aggregate demand from AD_0 to AD_1.

- This will cause a decrease – ultimately, remember that there are lags to take into consideration – in the demand for labour, shown by the decrease in the aggregate demand curve for labour from ADL_0 to ADL_1 in diagram III. This creates a cyclical addition to unemployment of $U_1 \Leftrightarrow FE$ – **cyclical unemployment** – increasing total unemployment from $FE \Leftrightarrow U^*$ to $U_1 \Leftrightarrow U^*$.

OUTSIDE THE BOX

The kinked supply curve for labour

The effect of wages being perfectly downward sticky has virtually the same effect as a **minimum wage**. (See Section 2.1 on minimum prices.) A decrease in the aggregate demand for labour from ADL_0 to ADL_1 does not clear the market since the real wage rate remains at W^*. In effect, this creates a **kinked supply curve** for labour.
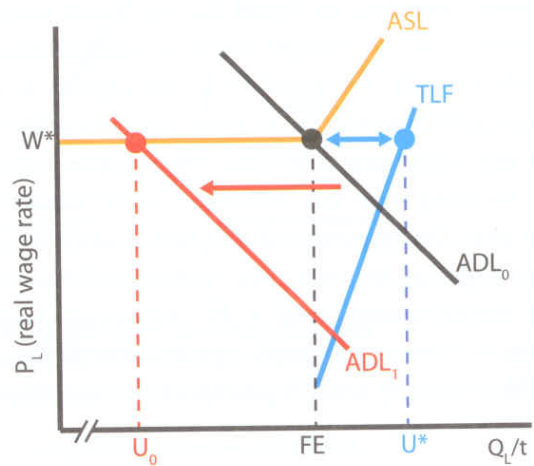
Real wage unemployment

By now you will need some revision. Here are three introductory points to real wage unemployment:

1. *Disequilibrium unemployment* – e.g. where unemployment exceeds the *equilibrium* or *natural* rate of unemployment – can be explained by a decrease in economic activity which leads to a decrease in the aggregate demand for labour. Since wages are 'downward sticky' the labour market does not clear and the addition to total unemployment is called cyclical or demand-deficient unemployment. This is the Keynesian/demand-side view of disequilibrium unemployment.

2. There is another way to view disequilibrium unemployment, namely that the labour market does not clear at the real wage rate because there are structural hindrances which disallow the market to clear. This is known as **real wage unemployment**. In short, the real wage rate is too high. The real wage is too high since there are 'labour market impurities' or 'rigidities' disallowing the market to clear. For example, minimum wage legislation and union bargaining power helps keep wages too high resulting in non-market clearing real wages. This is the new-classical view. Here's another clue; real wage unemployment is also known as **classical unemployment**. Classical theory views labour markets as not entirely dissimilar to the market for goods … such as, em, chickens.

Definition: 'Real wage unemployment' (or 'classical' unemployment)

Real wage unemployment is the new-classical view that any addition to equilibrium unemployment is due to labour market imperfections such as minimum wage, union bargaining power, high social and unemployment benefits and other labour market rigidities. These imperfections keep real wages too high and disallow market clearing, leading to increased unemployment.

The difference between cyclical and real wage unemployment is worlds apart – since the former is based on Keynesian premises and the latter on new-classical – but you'd have a hard time discovering these differences by only looking at the labour market diagrams. You see, the difference lies not so much in the fact that there *is* an excess of labour supply at a given wage rate but in *why* there is disequilibrium unemployment – and *what should be done* about it.

Figure 51.6 is virtually the same as in the diagram showing demand-deficient unemployment; real wage rates (W_0) are such that the market has not cleared, creating disequilibrium unemployment of $U_0 \Leftrightarrow U_1$. Why is there disequilibrium unemployment in the economy according to real wage theory?

$U_0 \Leftrightarrow U_1$: **Real wage unemployment** arises since at W_0, U_2 labourers are willing to work but only U_0 are demanded by firms. The real wage is too high and total unemployment is U_0 to U_1.
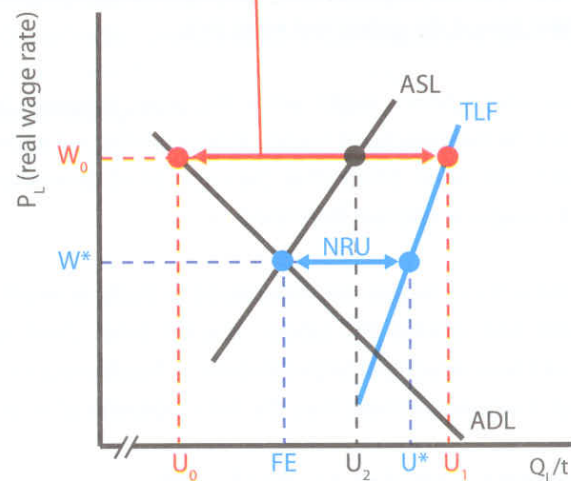

Figure 51.6 Disequilibrium unemployment – real wage unemployment

According to new-classical views, when the real wage rate is above market equilibrium wage, W^* in Figure 51.6, there will be more labourers willing to accept jobs than there is demand from firms. More labourers are willing to accept a job (e.g. there is a movement along the **ASL** curve) at W_0 but there is less demand from firms, shown by the quantity demanded for labour at U_0. There is now more labour willing to take jobs than there are offers of jobs. According to the new-classical view, *disequilibrium unemployment* (e.g. unemployment above the natural rate) exists because labour market forces have not been able to clear the market by lowering the real wage rate sufficiently. The wage rate of W_0 is above the market clearing rate of W^*, creating real wage unemployment of $U_0 \Leftrightarrow U_2$. Total unemployment is thus $U_0 \Leftrightarrow U_1$, above what would be the level of unemployment if the labour market cleared at a wage rate of W^* - the **natural rate of unemployment** of FE $\Leftrightarrow U^*$.

MACROECONOMICS

How is the real wage rate set at above market clearing level (W* in Figure 51.6) and why can it remain there?

The new-classical view gives three main reasons:

1. One main reason is that the government might have instigated *minimum wage legislation* which sets the price of labour higher than the labour market would.

2. Another common argument is that social welfare states might have *social/unemployment benefits* which would decrease the propensity of people in the labour force to accept wages below a certain rate. High unemployment benefits mean that the opportunity cost of remaining unemployed is relatively low – labourers lack incentives to take jobs at the going real wage rate.

3. *Trade union power* might serve the same purpose; by successful bargaining of wages. Wages are bid up above equilibrium level. (Note that the concept of downward sticky wages is relevant here too.)

All in all, these three factors are considered to *build in market imperfections* and hinder the labour market from clearing. The result of these structural impediments (= hindrances) is a higher level of unemployment than the full employment level.

Government policies to deal with unemployment

The two big policy areas are the Keynesian and new-classical approaches to disequilibrium unemployment and will be dealt with in further chapters. Here we will look at possible solutions to the three components of equilibrium unemployment; frictional, seasonal and structural.

Decreasing frictional unemployment

Frictional unemployment is conceptually based on the notion that a suitable job exists for the job-searcher and that it is a matter of finding that job. The basic issue in decreasing frictional (also known as 'search') unemployment is to shorten the time workers spend between jobs. That narrows down the options – in reality, it is a matter of creating a better 'match' between seekers and employers and/or incentivising workers to take jobs that might not be at the top of the list. Yes, we are back to the 'able and willing' concepts of supply and demand.

- Government run work offices with **centralised computer systems linking employers to job searchers**

have worked very well in Sweden, Denmark and Germany.

- **Worker programs** aimed at increasing skill levels, such as apprentice programs and government run **training schemes** have been implemented in many European countries.

- A particularly Scandinavian attempt at lowering search unemployment for young people was to offer firms a **'discount' on labour taxes** when employing under-20s. (Yes, this met with some considerable criticism and it was finally dropped when it turned out that firms' demand for over 20s labour *dropped* more than new hiring of youths increased!)

- **Lower income tax rates for low paying jobs** can increase the net wage and thus incentivise workers to take 'second best' options while continuing to search.

- Governments can also increase the opportunity costs of remaining unemployed during job searching by **lowering social and unemployment benefits**.

Decreasing seasonal unemployment

Here too is the issue of getting labourers to accept jobs – but often it means that labour has to move geographically and/or industrially. A lobsterman in the off-season will have to move elsewhere and an Alpine ski instructor might find herself having to work in a restaurant. It is worth noting that since seasons are rather regular and therefore predictable, one often finds that people in seasonal industries manage to earn enough over the course of the season to get by the rest of the year. For example, my favourite island in Greece, Sifnos, sees a population go from just over a thousand during the winter to well over 3,500 during the May to September high season. Barba Niko, my wonderful Syfnian 'uncle' who lovingly chastises me for going on long distance runs in 40C heat, works 16 hours a day but can then spend a rather chilled-out off-season in Athens.

- Frequently market forces seem to take care of many cases of seasonal unemployment. Many regions rely on an influx of temporary labour during peak times and revert to 'back burner' levels in the off season. Good examples of this are found in tourist areas (Greece), wine making regions (France) and extensive sheep farms (Australia).

- Central and regional government can offer various forms of *regional incentives* aimed at reallocating labour from areas with highly seasonal demand for labour to areas in need of labour, for example free train fares to job interviews and financial help in moving to another region.

Decreasing structural unemployment

Keep in mind that structural means 'build into the fabric of society', e.g. unemployment arising from a permanent decline in the demand for certain types of labour. The problem is like a very difficult form of frictional unemployment where workers need to find alternative jobs in other fields. However, the 'time between jobs' in the case of structural unemployment can be a very long time indeed since a coal miner (sunset industry) with 30 years experience losing his/her job will have a hard task getting a job as a computer programmer (so-called 'sun-rise industry'). Putting a harsh point on it, what needs to be done is *increase the adaptability of the workforce* so that time between jobs is cut down. Such policies, commonly referred to as **structural reforms**, take on both interventionist and market-based forms:

Interventionist policies

- Government-run retraining schemes and tax breaks to firms which supply re-education for redundant workers can increase labour market flexibility.

- Several Nordic countries (Sweden and Finland for example) have 'outsourced' government agencies to regions with high levels of structural unemployment. This also goes for universities and R&D institutions. Higher paid jobs create regional multiplicative effects and increase demand for labour in other areas such as services and recreation.

- Subsidies and grants can help firms establish in areas with sectoral/regional unemployment.

- .... other supply-side policies, see Chapters 60 – 62

Market-based policies

- Lowering unemployment and social benefits can create incentives for workers to move geographically and/or take jobs at lower wage levels. This was a policy during the Thatcher era in the 1980s.

- Labour market legislation making it easier to hire/fire can increase labour mobility between jobs. This was done in Denmark during the 1990s which has resulted in one of the world's most mobile labour force and also a consistently lower unemployment level than the rest of the EU.

- … other supply-side policies, see Chapters 60 – 62

The overall aim of market solutions to unemployment is to alleviate supply and demand mismatches, i.e. to improve labour allocation. Proponents of market solutions point to the US and Great Britain as examples of how long-run unemployment rates fell markedly during and after the Reagan and Thatcher reforms of the 1980s. This will be covered in depth in Chapters 60 – 62, supply-side policies.

MACROECONOMICS

Summary & revision

1. **Equilibrium unemployment** consists of structural, frictional and seasonal unemployment.

   a. **Structural** unemployment arises when there is a permanent decline in the demand for labour in certain industries. Three types of structural unemployment exist: *sectoral*, *regional* and *technological*.

   b. **Frictional unemployment** is the result of workers being in-between jobs – they are actively seeking jobs, which is why this type of unemployment is also called *search* unemployment.

   c. **Seasonal unemployment** arises for workers in off-season in seasonal industries such as tourism.

2. **Equilibrium unemployment** is also known as the **full employment** level of unemployment … which is also the **natural rate of unemployment**.

3. There are two main views of how **dis-equilibrium unemployment** arises:

   a. **Cyclical** (or demand-deficient) unemployment – a Keynesian view that the disequilibrium is caused by low aggregate demand.

   b. **Real wage** (or classical) unemployment – a new-classical view where unemployment above the natural rate of unemployment is caused by too-high real wages in the economy.

4. **Government policies** to reduce unemployment:

   a. *Frictional* – examples include improved matching of available labour and firms via centralised job search agencies; training and education; and lower income taxes on low-paying jobs.

   b. *Seasonal* – not a major policy issue, yet various regional incentives have often been implemented to reallocate labour from low to high demand areas.

   c. *Structural* – interventionist policies focus on use of government subsidies and tax breaks to increase demand for labour in regions and industries, while market-based policies are intended to increase labour mobility by incentivising workers to move and/or take available jobs and ease the hiring of labour for firms.

52. Low and Stable Inflation

Key concepts:

- Inflation, disinflation and deflation
- Measuring inflation using the consumer price index (CPI)
- Core – underlying – rate of inflation
- Weaknesses of inflation figures
    - o Consumption and quality bias
    - o Inflation is an average
- Producer price index
- Consequences of inflation

HL extensions:

- A weighted price index
- Calculating the price index (weighted data)

"Inflation is when you pay fifteen dollars for the ten-dollar haircut you used to get for five dollars when you had hair." (Sam Ewing, author)

Inflation, disinflation and deflation

Consider two possible events in an economy:

1. the price of agricultural goods rises due to a bad harvest; or

2. the price of all consumer goods rises due to an increase in VAT. Neither of these will technically constitute inflation since the change in the price level either only affects certain goods or occurs 'one-off'.

> **Definition: 'Inflation'**
>
> **Inflation** is defined as a consistent increase in the general (i.e. average) price level, measured by the consumer price index (CPI) or GDP deflator.

Inflation shows the change in the price level between two time periods; this is the *rate of inflation*. For example, a change in the consumer price index – CPI – from 108 to 112 means that the 'speed' of price increase is 112/108, i.e. a 3.7% rate of inflation. Frequently it is necessary to sub-divide the term into various types of inflation.

- **Creeping inflation** would be 'moderate' and involve a change in prices of a few percent per year. Most OECD countries showed inflation rates of this magnitude during 2003 – between 1% and 6%.[1]

- **Hyperinflation** is a situation where inflation hits triple-digits, for example when inflation in Brazil peaked at over 2500% during the early 1990s.[2]

This is what happens when 3 kgs of paper money buys one roll of toilet paper. Taken on the South African side of the Zimbabwean border at Beitbridge. The annual inflation rate in Zimbabwe

1   *OECD in figures*, 2003 edition, see www.oecd.org

2   IMF working paper, WP/01/50, *High inflation and Real Wages*, by Benedikt Braumann, 2001

MACROECONOMICS

in November 2008 was 89,700,000,000,000,000, 000,000%. I'm not making this up – heck, I can't even *pronounce* this figure! (From a Zimbabwean colleague.)

Deflation is strictly speaking the opposite; a persistent fall in the general price level. However, the term is often used a bit more loosely to convey downward pressure on prices due to a decline in economic activity, e.g. a 'deflationary gap'.

Disinflation A concept confusingly similar to that of deflation is **disinflation**, which is a fall in the *rate* of inflation, i.e. when inflation drops from say 7% per year to 3% per year. Disinflation occurs either when aggregate supply increases faster than aggregate demand – benign (= good) – or when aggregate demand falls faster than aggregate supply – malign (= bad) disinflation. An example of disinflation would be China during the latter 1990s and an example of deflation would be Japan (see below) during the same time period.

> **Definition: 'Deflation'**
> When the average price level, determined by the CPI or the GDP deflator, falls consistently during a time period, there is **deflation**, or negative inflation.
>
> **Definition: 'Disinflation'**
> When there is a fall in the rate of inflation, say from 5% to 4%, one speaks of **disinflation**.

Costs of deflation

"Inflation is taxation without legislation." Milton Friedman

Deflation can be good, bad, and pretty darned ugly. Economists commonly differentiate between two types:

- **Good, or benign, deflation:** This is caused by an *increase in aggregate supply*. Diagram I in figure 52.1 shows how the price level falls when short run aggregate supply outpaces demand; the price level falls from P_0 to P_1. Such deflation might result from increasing productivity and cannot be considered harmful since the economy is growing and real incomes are increasing. In reality, **disinflation** (falling inflation *rates*) has become the norm in industrialised countries, where average inflation was 5% during the

1980s but had fallen to around 2% by the end of the '90s.[3]

- **Bad, or malign deflation:** However, if the price level falls due to a *decrease in aggregate demand*, as in diagram II, Figure 52.1, there can be serious and long-lasting negative consequences for the economy; *malign* deflation. An economy experiencing a recessionary period that becomes protracted might cause households and firms to decrease consumption and investment to ride out the bad times and wait for the good times. This can actually prolong the recessionary period when households and firms decrease expenditure in favour of saving. Remember, a fall in the price level will increase the value of money. If households expect prices to continue to fall they will put off expensive purchases in order to get more for their money. This fall in aggregate demand can therefore confirm firms' beliefs that less investment is necessary which together with the decrease in consumption can become self-reinforcing in the economy.
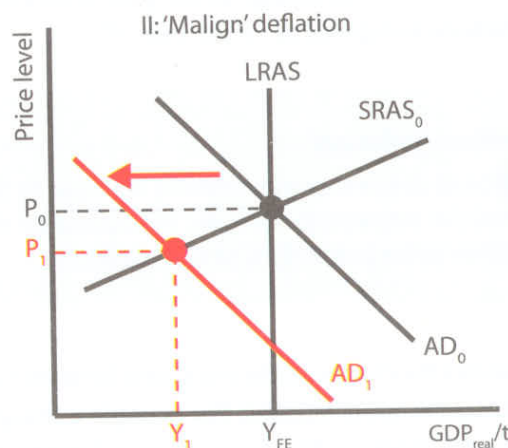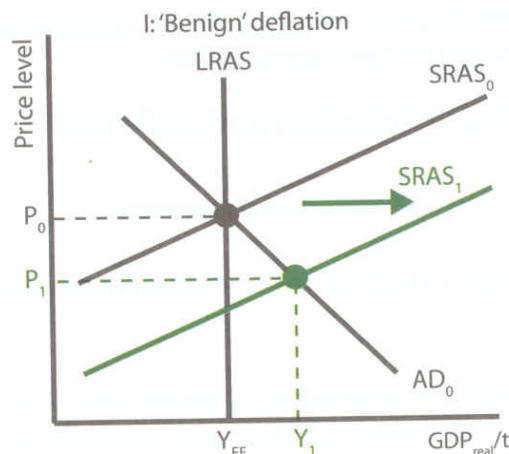

Figure 52.1 'Benign' and 'malign' deflation

---

3   *IHT, A new economic era – A global shift to deflation*, May 22, 2003

One might say that *malign deflation* cures inflation something like lung cancer cures smoking and I dare say that most economists would agree that deflation is a far greater threat to economic stability and growth than inflation. The self-reinforcing loop – known as a **deflationary spiral** – created by falling prices → expectations of falling prices → lower aggregate demand → falling prices … etc. is a most powerful force for fiscal and monetary policy to overcome. In fact, many textbooks use the Great Depression of the 1930s to describe the effects of continuously falling aggregate demand and resultant deflation. As prices, expenditure, output and incomes fall there will be increasing unemployment which further dampens aggregate demand and can quite possibly become permanent as some sectors fold and others see permanent reductions in demand. This might lead to a higher natural rate of unemployment.

It seems that deflation, once it becomes entrenched (= deeply rooted) cannot be dealt with easily. The 'trick' is to *create inflation by increasing the inflationary expectations* of households and firms, but the severity of the situation seems to resist standard monetary and fiscal policies. The solutions offered can therefore be at the extremist end of policy-making, where common suggestions involve some or all of the following: quickly lowering interest rates (see **liquidity trap** in Chapter 59) as soon as prices show a tendency to fall; very publicly announcing that the Central Bank has increased the target rate of inflation; large scale purchasing of bonds to create additional liquidity on the market; depreciating/devaluing the currency to increase exports; and printing 'funny money', i.e. printing consumption certificates which can only be used for consumption and then sending them to households in order to boost consumption and aggregate demand.[4]

Measuring inflation using the consumer price index (CPI)

The basis for deflating nominal values into real values is an index showing price changes. The most commonly used price index is the Consumer Price Index (CPI). This index series is arrived at by following an 'identical basket of goods' in a country over time in order to show the change in the overall price level. Figure 52.2 below follows such a basket in Ireland over a period of six years. The CPI, which is a stock concept, is measured in December each year. Its start date was 1995. This is the *base*

4 In fact, Milton Friedman's tongue-in-cheek suggestion for solving deflation was for the government to print money and fly around in helicopters and unload the bills on a happy citizenry – the bills would be time-limited in order to induce consumption rather than saving. This was actually attempted – without helicopters – during the deflationary crisis in Japan.

year, i.e. the year all coming index values (e.g. the price of the basket) will be compared with. The formula for concocting a price index is to divide the nominal price of the basket of goods with the price at base year values:

$$\text{CPI at } t_0 = \frac{P_{\text{basket}} \text{ in } t_n}{P_{\text{basket}} \text{ in } t_0} \times 100$$

Where t_n is the year being looked at and t_0 is the base year value, i.e. the original price of the basket.

This is usually where a few of my people start to freak out, thinking that it's all so *incredibly* complicated. Wrong. If *I* can get it, you can! Just imagine that a basket containing Guinness, mutton, crab paté and potatoes increases in price from €50 to €55. Now, slip these figures into the formula above. Clearly the price has increased by 10% or – indexed – from 100 to 110. Now add a few more goods to the basket ….

1995 (t_0)

20 pints of Guinness: €30

10 kg's of mutton: €35

10 cans of crab paté: € 40

100 kgs potatoes: €20

.

.

.

… and 1,000 other consumer goods

_____

Total cost of basket in Dec 1995 = €200,000

$$\text{CPI } (t_0) = \frac{€200,000}{€200,000} \times 100$$

$$\text{CPI}_{95} = 100$$

1996 (t_1)

20 pints of Guinness: €32

10 kg's of mutton: €36

10 cans of crab paté: € 41

100 kgs potatoes: €19

.

343

.

.

… and the same 1,000 goods as before

_____

Total cost of basket in Deck 1996 = €203,200

$$CPI\ (t_1) = \frac{€203,200}{€200,000} \times 100$$

$CPI_{96} = 101.6$

2001 (t_0)

20 pints of Guinness: €33

10 kg's of mutton: €39

10 cans of crab paté: € 42

100 kgs potatoes: €18

.

.

.

… and the same 1,000 goods as before

_____

Total cost of basket in Dec 2001 = €238,000

$$CPI\ (t_6) = \frac{€238,000}{€200,000} \times 100$$

$CPI_{01} = 119$

Figure 52.2 Calculating the CPI (un-weighted index)

The CPI is amassed by having a representative basket of consumer goods over time to show overall price changes in a country. Every country's 'basket' will have different goods in varying quantities, but the basic methodology is the same. The base year is 1995 in Figure 52.2, rendering a CPI value of 100. Summing the total cost of the basket at consecutive points in time shows that the price level has increased to 101.6 during the first year, or a 1.6% increase. Over the entire six year period the price level went from 100 to 119, a 19% increase in the price of the basket, e.g. the *average* price level has gone up by almost a fifth.

POP QUIZ 52.1

Consumer Price Index

1. What can you say about inflation by looking at the CPI above? Is the rate increasing or decreasing?

2. What has happened to the *real* price of Guinness during the six years?

3. What has happened to the *real* and *nominal* price of potatoes during the period?

Core – underlying – rate of inflation

Economists often find it useful to exclude goods which are notoriously volatile, such as energy/oil and food/agricultural goods, in which case one refers to **core inflation**. When the overall consumer price index is adjusted by removing goods which are highly volatile over time it is easier to follow the trend of inflation over longer time periods. Core inflation is a reasonable *indicator* of future inflation rates.

Weaknesses of inflation figures

Like all measurements we use in economics, weaknesses arise. Four major weaknesses rear their ugly heads from within the values we compile for inflation.

Inflation is an average

Perhaps the most important limitation of using the CPI to measure inflation is that the index values are **averages**. Different households will not be equally affected by inflation caused by large price increases in certain goods. For example, households on lower incomes will spend proportionately more on food than high income households, so an increase in foodstuffs will affect low income households more. Another weakness dealing with averages is that not all goods will show the same rate of price increase. Taken together, one could say that every household – indeed, every person – in the economy will have an individual rate of inflation.

Consumption and quality bias

Products improve over time – some of them quite considerably. The first hand-held computer I purchased in 1998 cost $US500, had a black and white screen and 2 megabytes of memory. My next one, in 2003, cost $US400, has 65,000 colours and close to 150 megabytes of memory – plus a number of features not even dreamt of when I bought the first one. The Samsung Galaxy S I bought in December 2010 … forget it; the information will be obsolete (= out of date) by the time this leaves the printer. Anyhow, this on-going process of continuous improvement and innovation often grossly overestimates inflation since the enhanced value of better products is not taken into account in the CPI.

Substitution bias

The CPI does not adequately take into account that consumers will substitute expensive goods with lower-priced alternatives over time. The time lag between when increasingly substituted goods are also taken out or replaced in the CPI basket will overestimate inflation.

Weight and content bias

Another way in which the CPI overestimates inflation due to time lags arises when goods become obsolete. The CPI basket needs to represent household spending and since consumption patterns change over time the contents in the basket need to change. If there is a lag in adjusting the basket to consumption patterns then the basket will incorrectly measure and weight items in the basket thereby skewing inflation measurements. In the other direction, if the goods and weights are changed often then the CPI might be comparing increasingly different baskets of goods over time, which would weaken the value of lengthy inflation series.

Producer price index

The producer price index (PPI) is a measure of the average price of factor inputs used by firms in an economy. The 'basket' thus includes raw material (for example oil), intermediary products (plastic, fibre and ink made from oil) and finished goods (plastic white board pens). While all three stages have their own price indexes, it is the final output price version of the PPI that economists tend to focus on. Literally thousands of individual producer price indexes are comprised for virtually every industry in the economy. The summary – which, like the CPI is weighted – becomes the PPI.

> **Definition: 'Producer price index – PPI'**
>
> The producer price index – PPI – is an index measuring the change in the price of factor inputs, intermediary products and final goods seen from the vantage point of producers (not consumers). It is comprised of the selling prices of goods used in thousands of industries in an economy. It is assumed to be an indicator of how future inflation will move.

The PPI has a base year and is put into year-on-year increases, just like the CPI. However, do not confuse 'final output prices' with consumer prices! The PPI measures what *sellers* are getting rather than what end consumers are *paying*. In effect, the PPI vs the CPI shows the ability of retailers to increase prices (the profit margin) over that of what wholesalers received for the goods. Also, retailers will be subject to a variety of further price determinants such as sales taxes, government subsidies and distribution costs.

The PPI should theoretically precede the CPI since any increases in input costs for producers will be passed on to consumers – and picked up by the CPI. One should therefore expect the CPI to lag the PPI – and this is exactly why a good many investors, government institutions and central banks keep more of an eye on the PPI than on the CPI. (See Chapter 59 and the issue of time lags in monetary policy.) The theory of a lag between changes in producer prices and consumer prices is appealing – studying producer price changes would give central banks a 'sneak preview' of coming inflationary attractions and allow the banks to plan for interest rate changes.
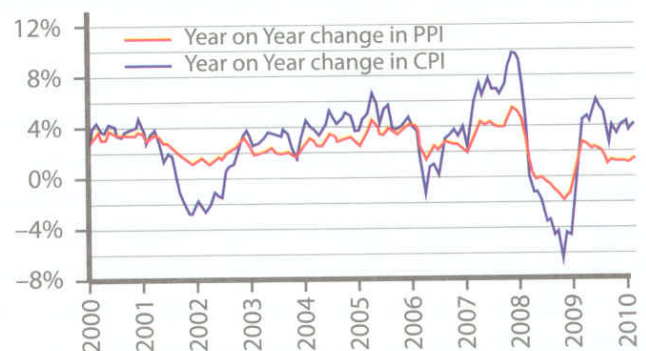

Figure 52.3 Correlation between the CPI and PPI, USA 2000 to 2010

MACROECONOMICS

However ...

the evidence for the *causal link* (refresh your memory with a few minutes of re-reading Chapter 3) is weak, mixed and inconclusive at best. Numerous studies have been conducted to find out a) whether there is correlation between the PPI and CPI (answer: yes, a regression coefficient of around 0.9 is not uncommon); and b) which causes which – the most common hypothesis is that changes in producer prices *causes* changes in consumer prices. There is no conclusive evidence across countries that consumer inflation is the result of producer inflation – and too many studies show the inverse! Do a simple test: take a ruler and put it vertically against Figure 52.3 and keeping the ruler perfectly vertical, run it slowly across the diagram from left to right. Now count how many times the red line (PPI) changes before the dark blue line (CPI). Yeah, it's pretty much 50/50 since in many cases the dark blue CPI curve changes *before* the red PPI curve, indicating instead that the causality goes from consumer prices to producer prices!

Consequences of inflation

The basic **function of money** is to provide the economic system with a *medium of exchange* in order to make transactions between actors on the market easier. Another key function of money is as a *store of value*, since it can be put aside and used later – which a bit more difficult with goods which have limited shelf lives such as carrots or piglets. This allows firms and households to *defer payments* over time. Finally, money is a most valuable *unit of account*, i.e. the value of goods and services can be put into a common unit so one doesn't have to continuously calculate the correct market ratio of carrots to piglets. Inflation – and especially hyperinflation – diminishes the functions of money and inflicts numerous costs on society.

Inflation decreases the ability of money to function as in 'money units per unit of output'. Modern economies are highly dependent on a functioning monetary system. Inflation has the effect of eroding the value of money, which can inflict serious damage on an economy if inflation is high enough. Yet too low an inflation level runs the risk of turning into deflation – which most economists consider a far worse a problem. Finding the 'Goldilocks rate'[5] of inflation is an almost magical feat.

Redistribution effects of inflation

Vulnerable groups such as pensioners, households dependent on social security benefits, stand to lose a great deal when inflation rates are high since they are often on **fixed incomes.** Even if these incomes are indexed to inflation rates, there will be time lags which will have adverse effects on purchasing power. In addition to this, workers in weak positions will not be able to bid up wages as much as workers in strong bargaining positions. (See 'In real life ...' story further on.)

Two other groups which stand to lose real income due to inflation are **lenders and savers**. Inflation has the tendency to decrease the distance between inflation and interest – the real interest rate – which will lower real returns for lenders. Similarly, savers will see how the real value increase in their savings will go down as it is eaten away by inflation and therefore diminishes future real purchasing power. On the other hand, borrowers will gain at the expense of savers and lenders, since inflation will also serve to erode the real debt of the original loan. Basically, income has been redistributed from lenders and savers to borrowers.[6]

Finally, the **financially strong** will suffer far less than the weak. The wealthy will have access to more and better information and thus the ability to cope with inflation by finding assets which are relatively secure in value-retention, such as land and other fixed assets.

Negative effects on growth

Inflation causes an increase in interest rates and will therefore have a negative effect on **investment and output**, both of which will adversely affect **employment rates**. Neo-classical and monetarist economists are quick to point to this particular effect of inflation.

Behavioural distortions in the economy

Inflation (and increasing rates of inflation) affects the 'view' of firms seeking to maximise returns on investment since the ratio between future profitability and the cost of loans is distorted when price levels increase. (Keynes, in keeping with his philosophical preference to be 'vaguely right rather than precisely wrong', referred to firms' expectations as the *'animal spirits'* of firms' investment plans.) Inflation also affects factor prices differently from final output prices, which adds further variables and thus complications to investment plans. Anything which causes **insecurity and lack of predictability** in firms' futures will serve to make them more cautious.

---

5 I.e. not too hot and not too cold. You know, the porridge ...

6 I know a good many people who bought houses in the early 1970s and are overjoyed at what high inflation rates did to their real debt over the next 20 years.

In addition to investment, consumption plans are often affected by inflation. Households have an incentive to increase consumption of durable goods if future inflation is expected to rise, which can actually create higher inflation rates. This is a macro force which to a certain extent can actually countermand falling investment in firms.

'Shoe leather' and 'menu' costs

Stable prices give the economy an element of transparency, in that firms and households can – within margins – safely foresee changes in prices and interest rates and plan accordingly. High inflation rates decrease the certainty in knowing correct (or at least 'fair') market prices for factors and goods, leading firms and households to spend more time searching for the best prices on the market. The time spent analysing the market yields opportunity costs in productivity; this is a 'shoe leather cost', i.e. the cost of walking around comparing prices. Higher inflation also means that businesses will have to continuously change their prices in order to keep up to date on the price level. Catalogues, price lists, price labels, vending machines etc will all have to be adjusted continuously in order to avoid losses in real terms. These 'menu costs' for firms can be quite considerable when inflation hits double digits.

Possible breakdown of the monetary economy

All the examples of how inflation damages the economy are in fact illustrations of how the functions of money wear down. In an economy with *hyperinflation* it is possible for the entire monetary system to collapse. For example, in Bolivia during 1985, the rate of inflation (at an annualised rate) soared to 11,750%![7] Now, ask yourself this; "Would I want to hold on to money when it will be worth 0.84% of its present value in one year's time?"[8] No, you will want to get rid of your Bolivianos as quickly as possible since every day you spend with cash in your pocket means an opportunity cost in terms of what it will buy you. The solution is to get something *tangible* for your cash such as consumer durables – or another currency. Since everyone else is thinking the same thing, inflation will be driven even further by increasing demand for goods but decreasing demand for holding on to the domestic currency. This self-reinforcing feedback loop results is a breakdown in the system since money becomes virtually worthless. People will spend enormous amounts of time tracking down the 'best' prices and ultimately simply resort to bartering.

This process is incredibly time-consuming and allocatively wasteful since **barter** involves high search costs for all parties in finding someone with *co-wants*; if you have piglets and need tomatoes then you have to find someone who not only has tomatoes but also wants a few piglets. This shows the importance of the transactions' function of money. The breakdown of the functions of money will cause massive disruptions to the government sector, since households and firms will have an incentive to put off tax-paying for as long as possible. This can cause severe cash flow problems for governments needing money with which to keep public services going.

Exchange rates, trade and inflation

Finally, inflation is strongly linked to the *external value of the currency*, i.e. the value of the domestic currency in terms of other currencies, which is the **exchange rate**. When domestic inflation is higher than in countries with which trade is done, there is a tendency for the value of the domestic currency to fall. The reason is that a higher domestic price level will make domestic goods less competitive internationally. As foreign demand for domestic goods falls, the derived demand for the domestic currency (which foreigners need to buy goods) will also fall – as will the price of the domestic currency, which is non other than the exchange rate. To use the Bolivian example again, the Bolivian currency (the peso, until 1987) had an exchange rate in 1980/'81 of 25 pesos to the US dollar – or 1 peso = 4 US cents. The official rate towards the end of the hyperinflation in 1985 was 75,000 Bolivian pesos per US dollar, while the black market value was over one million pesos per dollar; that's 0,0001 US cents per peso![9] (In other words, the *external value* of the Bolivian peso – measured in terms of the US dollar – went from 4 US cents to 0.0001 cent. The peso depreciated (= fell in value) by 99.9975%.)

As Chapters 67, 71 and 72 will show, the exchange rate in its turn is strongly linked to trade. A fall in the price of the domestic currency – *ceteris paribus* – will mean that domestic goods are relatively cheaper for foreigners to buy, since they will get more domestic currency for their foreign currency. Correspondingly, domestic consumers/firms will find foreign goods more expensive, so imports will fall. Thus, a fall in the exchange rate of the domestic currency can increase export revenue and decrease import expenditure.

---

7 Abel & Bernake, pages 459 – 461

8 The purchasing power of one Boliviano after one year of inflation at 11,750% is calculated as 1/[inflation/100]+1. At 50% inflation your Boliviano would be worth 1/[50/100]+1 = 0.666..., or 66.6% of the original value. At an inflation rate of 17,750 one Boliviano will be worth 1/[11,750/100]+1 = 0.0084, or 0.84%.

9 *Federal Research Division of the Library of Congress* at; http://countrystudies.us/bolivia/49.htm

Story Time!

In real life – Effects of inflation in Russia after 1991

One of my many amazing students is Miss Anna Semenikhina, from Russia. She was kind enough to interview her mother and grandparents for this little case study on the effects of rampant inflation.

Here's her mother's story word for word. 'Inflation happened really suddenly. It had not been 'announced' on TV, but things happening around you made you think of what had happened. Due to the situation in my family, related to my husband's sickness and then death, I moved to the evening department of the University and began working at one of the very first Russian 'co-operatives', which was a small privately owned company, with free commercial prices to its products. It was hardly possible to call it 'products', as there were not much to sell in the country. The company was selling those things, which had been forbidden, haunted to some degree, and lacked all sort of information about them. Wushu, yoga and Maharishi Aurveda with Transcendental Meditation as its part. I was one of those rare ones, whose income was correspondent to the prices in the grocery shops with free prices and co-operative shops, where you could buy everything. I believe in July 1990 my salary went up to 2000 roubles, at the same time most of the salaries varied from 90 to 200 roubles. One could buy a new 'Zhiguli' (LADA) car for 4000 roubles (with the right connections, of course). This is just an idea about how much money it was at that time.

In real life, the whole population became divided into governmentally owned workers and 'kooperativshiki' (co-operativshiks), people with money. The government didn't really understand how much money some people were making at that time, as all that had been going on for 3 to 5 months; and all those people 'up there' couldn't react to that and were not used to reacting. The first co-operative restaurant, which opened in Moscow, attracted the attention of the authorities and made them look at the situation differently due to the following.

The man, who owned the restaurant, was a Communist party member, and being an honest member of the party he paid party dues every month. I can't remember exact figures, but one month he paid dues equivalent to thousands of dollars. I will try to recall the prices in his restaurant. I believe I could eat there for 40 roubles. The governmental rate to a dollar was 0.35 (?) kopecks, (only used for non-currency operations for the diplomats and such), but black market rate was 13–16 roubles for a dollar. So it's not hard to calculate the real equivalent cost of food there.

The black market: the only real market in the country. To put it in two phrases, everyone in the whole country had certain connections to the people, which were selling stuff. The black market penetrated all spheres of life and existed as a parallel reality. It was very powerful. I can say it was a parallel economic system in the country. One of the most prestigious professions was a seller in a food store. All actual sales happened through the back door, while the counters offered matches, salt and canned fish. That was not risky at all. The real dangerous black market business was selling and buying hard currency. Article 88 of Russia's Criminal Code. Up to 8 years in prison.

Pensioners: My father served in the army, it paid rather well to be in the army, and by the time of default he had about 20,000 roubles on his account, with which he could buy several cars or live trouble-free for the rest of his life. I had a friend, who was deeply into black market and he started to advise my parents to take cash from the bank and start buying everything 'hard'. Gold, dollars, cars, whatever, things he could sell a bit later. My father, who served for his country all his life, wanted to believe that the country would not leave him in trouble and was waiting for the compensation to the last moment. He had never recovered from the betrayal. At the moment he died he was absolutely poor – with a pension equivalent to about 50–100 dollars per month.'

POP QUIZ 52.2

Inflationary Survival Quiz fpr Travellers

Picture yourself in Turkey: wonderfully hospitable people; fabulous food and drink; azure water at 27°C; some of the finest remnants of ancient civilizations in the world; and manufacturers of hand-woven oriental carpets of the highest quality. You spend a week in Turkey under the following premises; inflation is running at over 100% per year and the exchange rate for the Turkish lira is falling measurably and continuously every day.

1. How do you think restaurants and bars catering to tourists deal with the menu costs of changing all their prices daily?

2. There are foreign exchange offices everywhere. What is the most sensible way to manage your foreign exchange activities?

3. You buy a beautiful silk carpet for 2,500,000,000 lira – which is the equivalent of $2,000 at the time of purchase. Should you pay in cash or use a credit card?

4. Turkey is home to two of the ancient Seven Wonders of the World. Which? (This has absolutely nothing to do with economics but it's definitely cool.)

HL extensions

We have dealt with the basic methodology in measuring inflation in Section 3.1, Nominal and real national income. The Irish economy was used to build an example of a consumer price index (CPI) and a GDP deflator. The weakness in the method used was that both baskets of goods totally **disregarded the relative importance** of different goods. Some goods are more important than others, for example housing will be a major consumption item for households and electricity for firms. Knitting needles and paper clips won't. What one must do to make a price index more applicable is to weight the goods in the basket in a manner that reflects their relative importance; rent for housing should have a far heavier weight than knitting needles in households' expenditure patterns just as electricity over paper clips in firms.

A weighted price index

Figure 52.4 shows a weighted basket of consumer goods for Ireland. Note that the same basic methodology is applied when constructing a broader measure of inflation such as the GDP deflator.

Figure 52.4 Weighted price index for Ireland

Base year (t_0)	Un-weighted CPI (t_1)	% of income spent on good	(1) Weight	(2) ΔPrice	Affect on CPI (1)×(2)
20 pints of Guinness: €30	20 pints of Guinness: €32	20%	0.2	6.70%	1.34%
10 kg of mutton: €35	10 kg of mutton: €36	30%	0.3	2.90%	0.87%
10 cans of crab pâté: €40	10 cans of crab pâté: €45	10%	0.1	12.50%	1.25%
100 kg potatoes: €20	100 kg potatoes: €21	40%	0.4	5%	2.00%
Total cost of basket = €125	Total cost of basket in next time period = €134	100%	1		Total: 5.46%
$CPIt_0 = €\frac{125}{125} \times 100$	$CPIt_1 = €\frac{134}{125} \times 100$	Weighted CPI shows an increase in the price level from 100 to 105.46.			
$CPIt_0 = 100$	$CPIt_1 = 107.2$	$CPIt_1$(weighted) = 105.5			

Calculating the price index (weighted data)

By assigning weights to the various goods in a basket of consumption goods it is possible to get a better picture of how price changes affect households. In Figure 3.5.19 it is evident that the 12.5% price increase in crab pâté does not affect households as much as the 5% increase in potato prices. This is exactly the purpose of weighting an index, since all methods for measuring inflation are estimates aimed at showing how the average household (or firm when using GDP deflator) is affected.

In the example above, the inflation rate of 5.5% in the weighted index is more accurate than a rate of 7.2% as the case is in the un-weighted price index. Using weights to show which goods are more important as proportion of total spending more correctly illustrates the impact on households. All countries use a weighting system of some kind when calculating indices such as the CPI and the GDP deflator. What varies between countries are of course the weights; goose-down jacket will not have the same relative importance for Finns and Algerians, just as Mongolians will hardly have the same weight for Guinness as the Irish (but they might actually have similar weights for lamb and mutton).

Summary & revision

1. **Inflation** is a general and consistent rise in the average price level as measured by the CPI or the GDP deflator.

2. **Deflation** is a general and consistent fall in average prices as measured by the CPI or the GDP deflator.

3. Deflation can be **benign** (caused by an increase in AS whereby the price level falls but GDP increases) or **malign** (a decrease in AD causes both deflation and negative growth).

4. The **consumer price index** (CPI) is an index which shows the price of a basket of goods at the base period (year) and over consecutive periods (years). The difference in the CPI is inflation.

5. **Core inflation** is the CPI with highly volatile goods removed – these are usually oil/energy and food/agricultural goods. When the most volatile goods are removed, the index has a better predictive value for future inflation.

6. **Weaknesses of measurements** of inflation include

   a. The CPI and GDP deflator are both based on *average* price changes.

   b. *Consumption and quality bias* – the goods compared over time are often far superior to the 'same' ones in the basket 10 years ago.

   c. *Substitution bias* – lower priced goods often replace higher priced goods over time as a natural process.

   d. *Weight and content bias* – the CPI doesn't change at the same rate consumers' habits to and the lag makes inflation figures unreliable.

7. The **producer price index** (PPI) is a measurement of the change of prices in inputs used by firms. Raw materials, goods in process and finished wholesale goods are price indexed and computed. It is posited – though highly uncertain in many cases – that the PPI will precede the CPI and thus act as good lead indicator of future inflationary movements.

8. The **consequences of inflation** are:

   a. **Redistribution effects** – households on *fixed incomes* will lose more than those who can bid up their wages. *Lenders* (creditors, e.g. banks) will lose at the expense of *borrowers* (debtors, e.g. households) since inflation eats away at real debt and the real value of money repaid to banks. The financially strong will lose more than the lower income groups.

   b. **Negative influence on growth** – higher inflation makes both households and firms reluctant to spend/invest due to the *insecurity* of future prices.

   c. **Shoe leather costs** – the increased time and resources spent on finding the lowest prices.

   d. **Menu costs** – firms will need to change prices frequently to keep up with inflation.

   e. Possible **breakdown of monetary economy** – hyperinflation can lead to widespread abandonment of money and the emergence of a barter economy.

9. **External balance effects:**

   a. Higher inflation (= internal value of money) means that the price of exports rises (ceteris paribus) and the price of imports falls. This leads to decreased exports and increased imports – a **negative effect on the trade balance**.

   b. Following the above, a decrease in exports will mean lower demand for the Home currency which means a **depreciation** of the home currency.

HL extension

10. A **weighted price index** sets values for goods indicating the relative importance of the good based on how households spend their money. For example, electricity will have a heavier weight than paperclips and using these weights to compute a CPI will better represent the effect on households.

MACROECONOMICS

53. Causes and Consequences of Inflation and Deflation

Key concepts:

- Causes of inflation
  - o Demand pull inflation
  - o Cost-push inflation
  - o Excess money supply
- Causes and consequences of deflation
- Counter-inflationary policies
- Evaluation of counter-inflationary policies

MACROECONOMICS

"Having a little inflation is like being a little pregnant – inflation feeds on itself and quickly passes the 'little' mark." Dian Cohen.
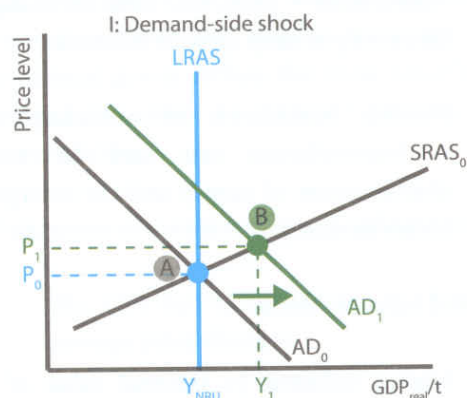
Causes of inflation

Two main causes of inflation are Keynesian in origin; *cost push* inflation arising from higher factor costs to firms, and *demand-pull* inflation which arises when aggregate demand in the economy outstrips available aggregate supply. A third, neo-classical/monetarist view, posits that inflation is demand-pull in nature, but that it is the underlying variable of *increased money supply* which is the root cause.

Demand pull inflation

Aggregate demand might rise for a number of reasons; stimulatory monetary/fiscal policies, greater consumer confidence, 'animal spirits' of firms … etc.

- **Demand-side shock, A to B:** When aggregate demand increases swiftly in the short run due to, say, greater demand for exports, there will be a *demand-side shock* and concomitant (= associated) increase in the price level, shown in Figure 53.1 (diagram I) as the shift form AD_0 to AD_1. [**And again, this is a perfectly acceptable way to illustrate demand-pull inflation!**]


I: Demand-side shock

$A \Rightarrow B$; the increase in demand for exports increases AD, causing a **demand-side shock**.

II: Demand-pull inflation
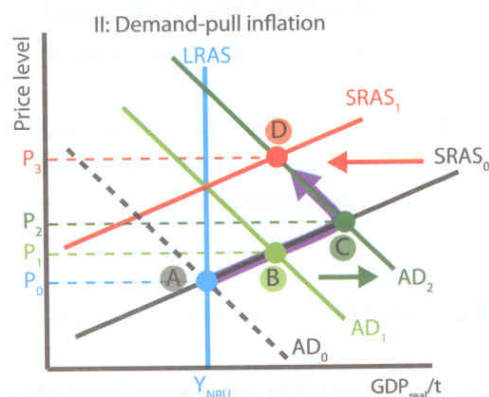
$B \Rightarrow C \Rightarrow D$; anticipation of higher inflation increases AD further (AD_1 to AD_2) and the price level increases to P_2. The subsequent fall in real wages causes a bidding-up of wages and as costs rise for firms, SRAS decreases from $SRAS_0$ to $SRAS_1$ and the price level continues rising to P_3. This is a round of **demand-pull inflation**.

Figure 53.1 I & II Demand-side shock, demand-pull inflation and demand-pull spiral

Figure 53.1 III Demand-side shock, demand-pull inflation and demand-pull spiral
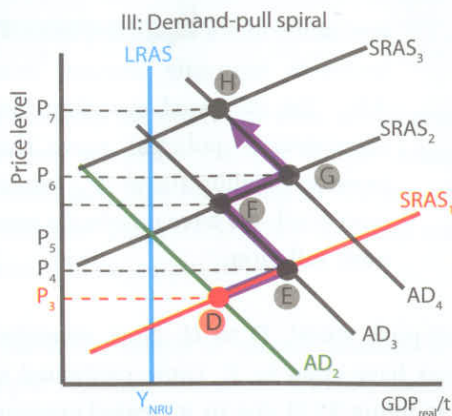
D ⇒ E ...⇒ H; if households again anticipate higher inflation at point D, then AD might again increase (AD₂ to AD₃). This leads to another demand-pull inflationary round and a **demand-pull spiral**; high inflation expectations → ↑AD → higher costs for firms → ↓ SRAS → higher price level, etc.

- **Demand-pull inflation, A to D:** Once again, the initial increase in the price level (P_0 to P_1) due to increased aggregate demand does not mean inflation *per definition*, but might rather set the stage for demand-pull inflation.
  - o Diagram II in Figure 53.1 shows how *inflationary expectations* cause aggregate demand to feed on itself, as firms' and households' spending plans increase in *anticipation* of higher future prices. Aggregate demand increases further, from AD_1 to AD_2, and the price level increases to P_2.
  - o This is untenable (= unsustainable) in the long run, since higher final prices cause labourers to suffer real wage loss, and subsequently they start to bid up wages.
  - o This results in higher labour costs for firms and a decrease in aggregate supply from $SRAS_0$ to $SRAS_1$. The economy has moved towards long run equilibrium (Y_{NRU}) but at a higher price level, P_3. The original shift in AD sets off a round of **demand-pull inflation** where aggregate demand increases beyond long run potential output.

- **Demand pull spiral, D to H:** If aggregate demand continues to rise – due to continued expectations of high inflation – then the economy can expect a process where prices increase → labour adjusts by bidding up wages → and firms scale back on production due to

lower margins between input prices (e.g. labour costs) and final output prices. This is a demand-pull spiral, as illustrated in Figure 53.1, diagram III.
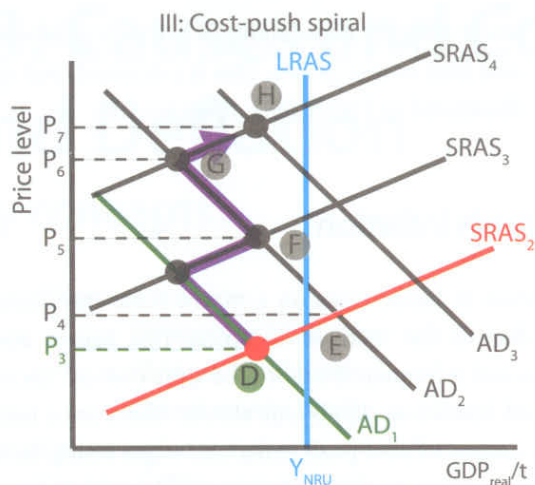
Cost-push inflation

An increase in factor costs on a macro scale can cause prices to rise due to the increase in costs. This causes **cost-push inflation** and is frequently associated with 'one-off' increases in price level, known as *supply-side shocks*. There are a number of possible causes of cost-push inflation; *wages* rising faster than productivity gains in the economy; a fall in the *exchange rate* driving up the price of imported raw materials and components; or an increase in *factor prices*, say the price of oil. All of these will shift aggregate supply to the left, shown in diagram I, figure 53.2.
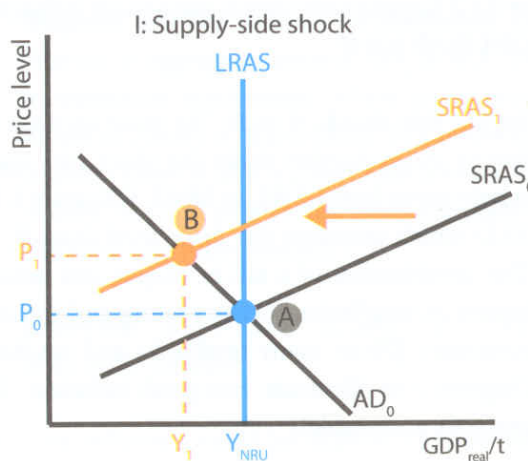
A ⇒ B; an increase in factor costs causes a decrease in SRAS. This is a **supply-side shock**.
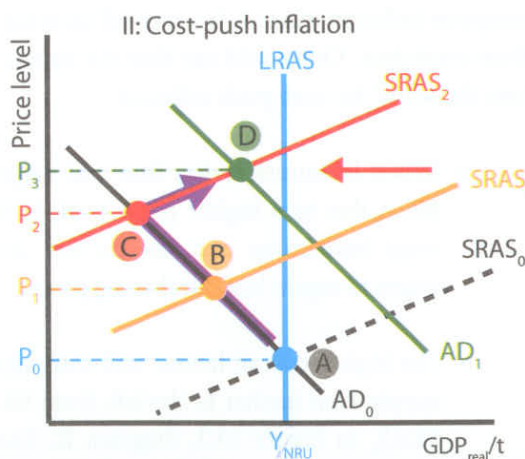
B ⇒ C ⇒ D; the decrease in real wages causes labourers to bid up wages and SRAS continues to decrease (SRAS₁ to SRAS₂). When higher wages fuel consumption, AD increases from AD₀ to AD₁. We've now seen a complete round of **cost-push inflation**.

III: Cost-push spiral

D ⇒ E …⇒ H; Should labourers continue to bid up wages after the increase in AD (point D), another round of cost-push inflation commences. We now have a **cost-push spiral**: ↑price level → ↑wages → ↑ costs to firms →

Figure 53.2 Supply-side shock, cost-push inflation and cost-push spiral

- **Supply-side shock, A to B**: The price increase in key factors of production shifts the short run aggregate supply curve from $SRAS_0$ to $SRAS_1$ (Diagram I, Figure 53.2) which increases the price level from P_0 to P_1. This combination of a fall in output and inflation is known as **stagflation** (i.e. *stag*nating and in*flation*ary economy). [**Note: many textbooks and teachers use** *Diagram I* **to illustrate cost-push inflation. This is perfectly acceptable!**]

- **Cost-push inflation, A to D**: Technically speaking, the supply-side shock illustrated in Diagram I does not comprise inflation, since it is a one-off increase rather than consistent. One could say that the supply shock 'sets the scene' for cost-push inflation:

  o When labourers realise that real wages have fallen due to a higher price level, individual wage bargaining and unions will *drive up wages* to regain lost purchasing power.

  o The higher cost of labour will shift aggregate supply even further to the left, from $SRAS_1$ to $SRAS_2$ in Figure 53.2, diagram II. The price level rises from P_1 to P_2.

  o Ultimately, the increase in wages (perhaps accompanied by expansionary policies in order to countermand increased

unemployment) increases consumption and increases aggregate demand from AD_0 to AD_1. The increased spending (and possible expansionary policies) move the economy towards equilibrium at Y_{FE} but at a higher price level. We have now had a round of **cost-push inflation**.

- **Cost-push spiral, D to H**: Now, consider that final prices have risen to P_3 (now continued in diagram III in figure 53.2) due to increased consumption and fiscal stimuli. Real wages have once again fallen due to the effects of inflation. Another period of bidding-up wages could lead to a further decrease in aggregate supply, shown by the shift from $SRAS_2$ to $SRAS_3$, and another round of cost-push inflation. Diagram III shows how successive shifts in SRAS and AD create an upward spiral, known as a **cost-push spiral** or **wage-price spiral**. The basic effect is; ↑price level → ↑wages → ↑ costs to firms → ↑ price level …etc.

The (in-) famous 1970s cost-push spiral

The most (in-) famous of supply-side shocks occurred during 1973 and '74 when OPEC managed to force oil prices upwards by 300%, the first **oil crisis**. Oil is vital to production since it provides energy, transportation, compounds for plastics etc, and when the price of oil quadrupled, firms' costs increased greatly, causing a severe supply shock and stagflation in most industrialised countries. When stagflation hit industrial countries, many responded by stimulatory policies which drove prices higher and resulted in cost-push inflation. Many countries saw repeated rounds of higher costs and wages which led to cost-push spirals.

The effect on the global economy was tremendous, with falling output levels and rising unemployment in most of the world. It also spelled the end of *pure* Keynesian demand-side policies; see long run Phillips curve following. As a most illuminating final footnote; recently declassified documents (December 2003) in Great Britain paint a most illustrative picture of just how serious the oil crisis was regarded by the President of the US in 1973/'74, Richard Nixon. In a copy of a report sent to British Prime Minister Edward Heath (dated 12th December 1973), Nixon put forward serious plans for the military occupation of Saudi Arabia, Kuwait and Abu Dhabi.[1] Perhaps the world is lucky that Nixon got his hands tied by the Watergate scandal around the same time.

---

1   *IHT*, January 2 2004, page 5; *US considered seizing Arab oil fields.*

MACROECONOMICS

Excess money supply

Assume an economy with only one firm, which produces 1,000 Widgets per year at a price of $1 per Widget. Additionally, the supply of money is $1,000. Assuming all Widgets are sold, what is national income? Easy; it's $1,000 – the same amount spent (E) and earned (Y) on output (O).

- What if the supply of money doubled to $2,000 but output remained the same? Quickly check Section 3.1 (Nominal and real income) to realise that nominal national income will increase but *real income will remain the same*. Assuming that the additional $1,000 flows around the economy at the same rate as the original $1,000, we can also fairly safely assume that the price of Widgets goes from $1 to $2 – the price level has doubled.[2]

- However, what if the ability of the firm to produce Widgets also increases during the time period, say that new technology and production methods enables the firm to produce twice as many Widgets at the same cost as before? Assuming that all are sold and that, once again, people use their notes and coins at the same rate as before, the price of Widgets will remain at $1, and real income will double to $2,000.

The above is a highly simplified and of course exaggerated example of how excess money supply causes demand-pull inflation.[3] Neo-classical/ monetarist economists view inflation as primarily *demand-driven* and caused by an increase in the supply of money (i.e. monetary growth) above and beyond the long run ability of the economy to increase the supply of goods – LRAS at the natural rate of unemployment (NRU). In simple terms; **if the increase in the supply of money is above productivity gains the result will be inflation**. The mechanism herein can be explained by the *monetary transmission mechanism*.

The monetary transmission mechanism

Monetarists view inflation as primarily caused by demand-pull, often phrased as 'too much money chasing too few goods'. One of the pillars of monetary theory is the overriding importance attached to the supply of money in the economy.
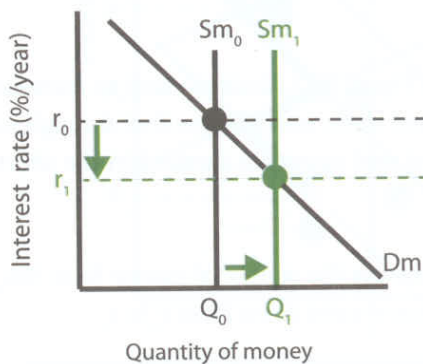
Monetarists view the demand for money as relatively inelastic, so an increase in the supply of money will have a large impact on firms' investment and households' consumption via falling interest rates.

The series of diagrams in Figure 53.3 illustrates how an increase in the supply of money and lower interest rates feed through to an increase in investment and aggregate demand in the short run.
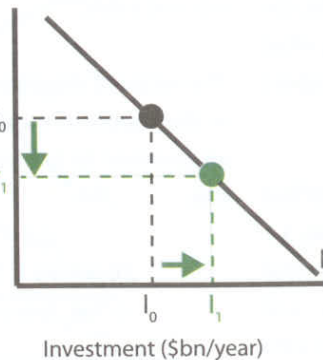
- **Diagram I**: Increasing the supply of money from Sm_0 to Sm_1 forces interest rates down from r_0 to r_1.

- **Diagram II**: Lower interest rates induce an increase in investment, which is shown by the movement along the investment schedule. Investment increases from I_0 to I_1.

- **Diagram III**: The increase in investment causes an increase in aggregate demand; AD_0 to AD_1. This is the **transmission mechanism**, where an increase in the supply of money is 'transmitted' to an increase in real output. (There is also a direct link between the lower interest rate and greater *household expenditure* which also fuels aggregate demand.)

- In summa: $\uparrow\Delta S_m \rightarrow \downarrow\Delta r \rightarrow \uparrow\Delta I$ and $\uparrow\Delta C \rightarrow \uparrow AD$ and $\uparrow\Delta$ price level (inflation)

The resulting inflation caused by increased money supply can thus be viewed as '... too much money chasing too few goods'.

---

2   Dear colleagues, I am aware that I am cutting a few corners in the cornerstones of the quantity theory of money and the Fischer equation.

3   For more depth on this – outside the syllabus (and this book) – look up 'quantity theory of money' and/or 'the Fisher equation'.
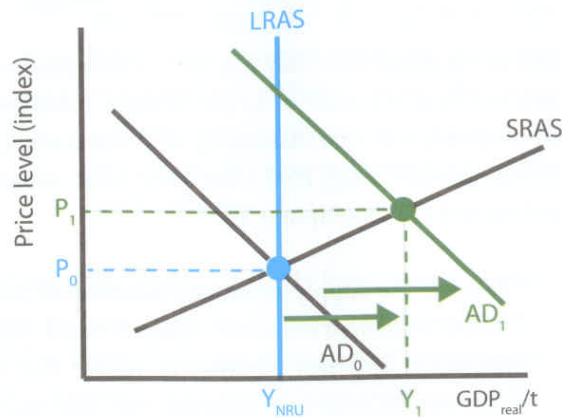
I: Supply and demand for money   II: The investment schedule   III: Affect on aggregate demand

Increased supply of money ..... ⇒ ..... lowers the rate of interest ..... ⇒ ..... which stimulates C and I in AD.

Figure 53.3 The (monetary) transmission mechanism

The neo-classical view of money-driven inflation – LR effects

It is important to realise that monetarist theory views any increase in money that is *not matched by an increase in real potential output* (LRAS) as being solely inflationary in the long run. As the increase in AD (Figure 53.3) has pushed equilibrium output beyond LRAS, the increase in real GDP shown in Diagram III will not last, since wages will rise to match labour demand, serving to increase costs for firms and thus push the AS curve to the left ($SRAS_0$ to $SRAS_1$) as in Figure 53.4. Output returns to the natural level of output, Y_{NRU} but at a higher price level.

Milton Friedman coined a main monetarist article of faith by stating that "Inflation is always and everywhere a monetary phenomenon".[4]
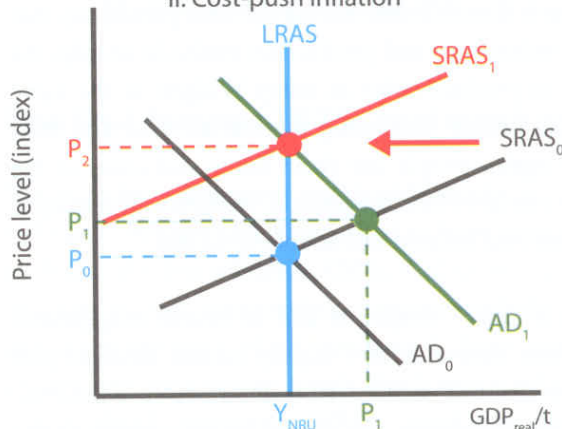

Figure 53.4 The (monetary) transmission mechanism

While I am oversimplifying, the basic idea of steering the economy primarily by regulating the money supply is in keeping with neo-classical theory since it advocates minimum government intervention in favour of a few simple rules and guidelines such as the Central Bank openly setting target rates of monetary growth and allowing the economy to adapt in terms of inflation and output. (See Chapter 59 for some depth on monetary policies available to the central bank.) Alternatively, the Bank can set inflation ceilings and then adjust money supply to keep inflation under control. A common 'rule of thumb' – simple but quite effective in fact – is that monetary policy should be tightened when nominal interest rates are lower than nominal GDP growth and vice versa.

---

4   One of my many cheeky students, Ms O'Connor, once sent me a picture of two machines; one had numerous dials, knobs, levers, wires, connections, scales and indicators on it while the other machine had a single button, 'On-Off'. The first machine was labelled 'Woman' and the second 'Man'. I often think of monetarist policy as a 'Man' machine, since all the various forms of fiscal policies are basically replaced with a single knob labelled 'Money: More ⇔ Less'.

POP QUIZ 53.1

Causes of Inflation

1. Why might governments in high inflation countries actually gain from the effects of inflation? Hint; governments issue bills and bonds to borrow money.

2. An economy 'imports' inflation due to the increase in price of irreplaceable raw materials. Unions push for higher wages to counter the cost of living increase and … now what? Use a diagram to explore possible outcomes.

3. Over a five year period, the following changes are noted in an economy:

Time period	Increase in money supply	Inflation	Real GDP growth
t_1	7.0%	6%	4%
t_2	10.5%	6.5%	2.8%
t_3	18.4%	8.5%	1.4%
t_4	9.6%	11.5%	-0.8%
t_5	9.0%	10.5%	-1.5%

Use the viewpoint of a neo-classical/monetarist economist to explain the pattern of inflation rates and growth during the five year period.

Causes and consequences of deflation

"Inflation is taxation without legislation." Milton Friedman

Deflation can be good, bad, and pretty darned ugly. Economists commonly differentiate between two types:

- **Good, or benign, deflation:** This is caused by an *increase in aggregate supply. Diagram I* in *Figure 52.1* shows how the price level falls when short run aggregate supply outpaces demand; the price level falls from P_0 to P_1. Such deflation might result from increasing productivity and cannot be considered harmful since the economy is growing and real incomes are increasing. In reality, **disinflation** (falling inflation *rates*) has become the norm in industrialised countries, where average inflation was 5% during the

1980s but had fallen to around 2% by the end of the '90s.[5]

- **Bad, or malign deflation:** However, if the price level falls due to a *decrease in aggregate demand*, as in *Diagram II, Figure 52.1*, there can be serious and long-lasting negative consequences for the economy; *malign* deflation. An economy experiencing a recessionary period that becomes protracted might cause households and firms to decrease consumption and investment to ride out the bad times and wait for the good times. This can actually prolong the recessionary period when households and firms decrease expenditure in favour of saving. Remember, a fall in the price level will increase the value of money. If households expect prices to continue to fall they will put off expensive purchases in order to get more for their money. This fall in aggregate demand can therefore confirm firms' beliefs that less investment is necessary which together with the decrease in consumption can become self-reinforcing in the economy.
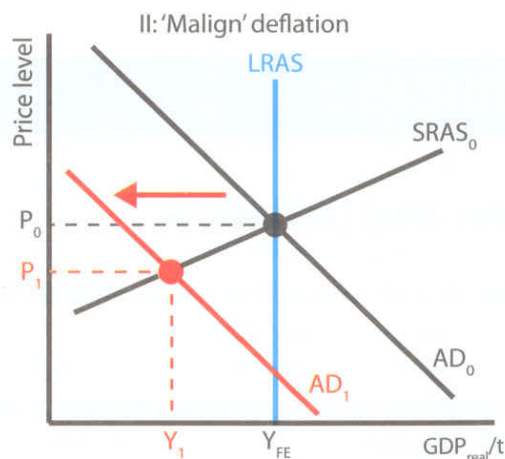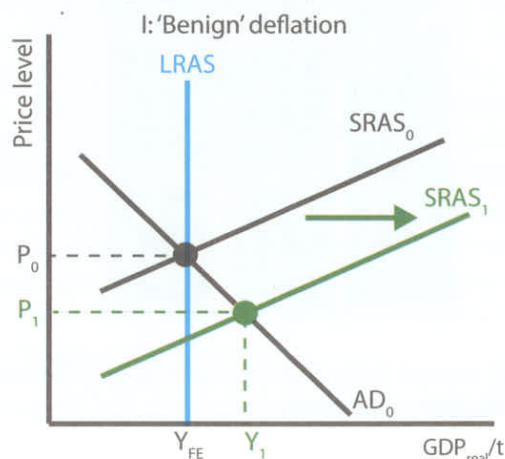

Figure 53.5 'Benign' and 'malign' deflation

5 IHT, *A new economic era – A global shift to deflation*, May 22, 2003

MACROECONOMICS

Postage stamps from the German hyper-inflation era of the inter-war years

POP QUIZ 53.2

Inflation and Deflation

1. Explain how an economist would regard an increase in cigarette taxes in terms of inflation.

2. How might the functions of money be put out of order during inflation?

3. Why might you, the student, stand to lose due to inflation?

4. Using a supply and demand diagram for your home currency, explain why double-digit inflation will have a negative effect on your exchange rate – i.e. cause your currency to fall in value.

5. Why might deflation in fact be considered 'good' for an economy?

6. Outside the box question: Draw a demand curve for investment (investment schedule) showing how the quantity of investment is totally unaffected by a decrease in the interest rate below 0.5%.

Case Study
Deflation – the ten-year sunset in Japan

The Land of The Rising Sun, Japan, was the envy of an astonished world during the 1960s, 1970s and 1980s due to its fantastic growth rates and low inflation and unemployment. GDP grew at an average of over 6% during the period 1960 to 1990 with unemployment rates hovering around 2%. Yet during the beginning of the 1990s, Japan was to become a case study in the difficulties in getting out of malignant deflation. The background is to be found in the property price bubble of the late 1980s where property speculation and loose monetary policy led to alarming levels of property prices. (For example, the classic story is that the land on which the Imperial Palace in Tokyo stood was at one point valued at more than the entire American state of California!).

When property prices started falling in 1990 a goodly proportion of the loans taken by speculators could not be serviced, and this created a severe lack of loanable funds – which in turn had a negative effect on investment.

In spite of falling interest rates, Japan has experienced falling prices, e.g. deflation, and stagnant or falling GDP during most of the period from 1995 to 2003.

The Central Bank, the Bank of Japan, loosened monetary policy to the point where interest rates were virtually zero during the period 1997 to 2003. This is explained by some – but definitely not all – economists as being a classic example of the liquidity trap. When interest is zero and the economy continues to shrink, the Central Bank basically 'runs out of ammunition' and there are few monetary policy tools left since it is impossible to lower interest rates and stimulate the economy. In addition, recall that real interest is nominal interest minus inflation. Thus, if nominal interest is zero and inflation is below zero, there is a negative rate of real interest. The repercussions are as astounding as they are simple:

- Negative real interest means that money increases in value even when you stick it in your mattress (or futon, in Japan). For example, Japan had inflation of –1% (or deflation of 1%) during 2002 while nominal interest was zero. Thus sticking money in a futon actually meant a real interest of 1% over a year. Negative real interest is therefore a disincentive to spend money.

- Monetary policy becomes ineffective since it is impossible to lower interest rates below zero. This would mean that banks would pay borrowers to take on debt!

- Corporate and household debt rises when prices fall, since a fall in prices increasesthe negative real interest rate. This is a major disincentive to firms and households in taking on debt to fund investment and consumption.

In labour markets, firms will need to increase productivity or cut wage costs in order to stay competitive, since falling final prices mean that profit margins are squeezed. In reality, firms will not be able to lower wages and will resort to laying-off workers. The reason is that when inflation is 6% a firm can accomplish real wage cuts of 4% by offering an increase of 2% nominal.

To achieve the same effect when inflation is zero the firm would have to cut nominal wages by 4%! This is not practically feasible in most countries, wherefore deflationary pressure inevitably leads to sizable lay-offs.

During the 1990s, the Japanese government also lowered taxes and increased government spending to the point where government debt exceeded national income by some 60% – the highest ratio of government debt to GDP in the industrialised world. The limits of fiscal policy in Japan show how the propensity to save can countermand a possible multiplicative effect of fiscal policy. The Japanese are a most frugal and careful people when it comes to saving; for most of the 1970s and 1980s the Japanese saved 15% to 20% of disposable income – twice that of the US. While the savings rate fell to below 10% during the 1990s, it was still high enough to have a severely limiting effect on tax reductions and increased government spending. In fact, in a fit of desperation the government issued time-limited 'spending certificates' valued at over USD 5 billion in 1998 to induce some spending and inflationary pressure in the economy. The result? Households used the certificates to replace money-based consumption and save more! In other words, there was little or no impact on aggregate demand. Neither fiscal nor monetary policy has been able to draw Japan out of what in effect seems to be a deflationary spiral.

Counter-inflationary policies

There is no 'cure all' for inflation and the policies applied will depend on the *type* of inflation and the *costs* government is willing to incur. The straightforward way to deal with inflation is to use demand-side policies (adjusting AD) for demand-pull inflation and supply-side policies (shifting AS) for cost-push inflation. Both have disadvantages, weaknesses and, as mentioned above, costs.

Using *contractionary* demand-side policies to combat inflation entails implementing policies which decrease aggregate demand. Four key methods of dampening aggregate demand are:

MACROECONOMICS

- **Increasing the rate of interest**: this monetary policy lowers consumption and investment and thus decreases AD
- **Decreasing the supply of money**: another monetary policy – this increases interest rates … etc …
- **Increasing taxes**: the fiscal policy of higher income taxes will lower consumption, higher profit taxes will decrease investment and both have a contractionary effect on AD
- **Decreasing government spending**: the flip side of taxes, the fiscal policy of decreased government spending has a direct effect in lowering AD

Figure 53.6 A shows how contractionary policies affect aggregate demand. Assuming general equilibrium at Y_{NRU} and the average price level indexed at 100, AD is increasing at a rate indicating AD_1 and inflation of 5%. One or several contractionary policies are implement and AD settles instead at AD_2, Y_2 and an inflation rate of 3%.
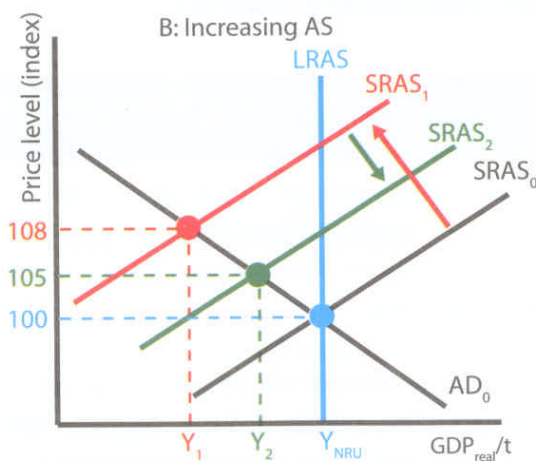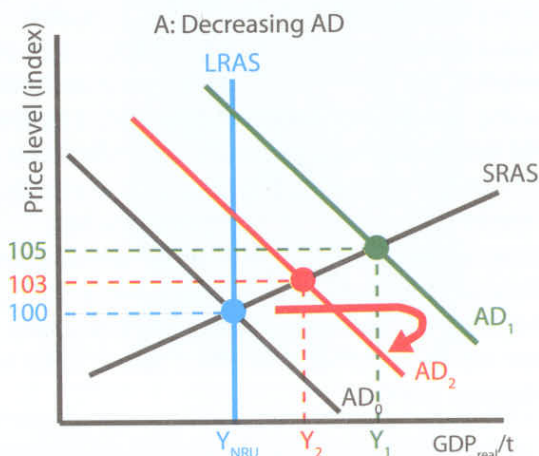

Using supply-side policies to lower cost-push inflation entails implementing policies which increase aggregate supply. Common examples of supply-side policies are:

- **Lowering marginal income tax rates**: higher net disposable incomes entices people to work more
- **Labour market policy changes**: legislation enabling easier hire rules for firms or decreasing minimum wage
- **Incentives for capital formation**: government can induce increased investment expenditure by giving firms tax breaks – e.g. allowing firms to deduct portions of investment spending from taxable profits
- **Interventionist policies**: government skills and re-training workshops can improve labour quality and also create better matches between labour supply and demand

Figure 53.6 B illustrates how $SRAS_1$ results in cost-push inflation of 8%. Over time – yes, this is a rather serious weakness – supply-side policies can shift SRAS back towards general equilibrium ($SRAS_1$ to $SRAS_2$) and a price level of 105 rather than 108.

Evaluation of counter-inflationary policies

This is a major issue and will be dealt with in greater depth in Sections 2.4 – 2.6. I limit the discussion here to the key points arising from the examples and Diagrams used in Figure 53.6.

1. To start with, contractionary demand-side policies can be hugely **unpopular with citizens** and are basically not much of a crowd pleaser for politicians hoping to get re-elected. President Nixon was most reluctant to implement much-needed contractionary policies when inflation started rising in the late 1960s.[6]

2. Contractionary policies often lead to increased **unemployment** and **lower incomes**.

3. There is the very real risk that contractionary policies are **too severe** and that the result is recession.

4. Any policy used will be subject to **time lags** – it can take up to two years for the full effect of interest rate changes to feed through in an economy. Fiscal policies can take

Figure 53.6 Inflation and demand-side and supply-side policies

6   See some fascinating economic history at Professor Bradford DeLongs site: http://econ161.berkeley.edu/econ_articles/theinflationofthes.html

even longer. There is therefore a very real risk that the economy is already cooling down when the policies kick in, which could in fact *worsen* the economic downturn.

5. Supply-side policies avoid the issue of decreased incomes but are generally **long term** solutions and will not have any immediate effect on inflation. Instead the aim is more along the lines of allowing for long term growth while limiting inflationary pressure.

6. Lowering income tax rates can have serious repercussions on governments' ability to **balance the budget**. (See Chapter 56.)

7. In reality the effect on the labour market of **lower personal income tax cuts is very limited** and simply does not increase labour supply.

Summary & revision

1. Economics identifies *three causes of inflation*:

   a. **Demand-pull inflation** caused by increasing AD

   b. **Cost-push inflation** caused by decreasing AS

   c. **Excess money supply** leading to demand-pull inflation

2. The **monetary transmission mechanism** describes how a change in money supply 'feeds through' to a change in AD. A $\downarrow \Delta Sm \rightarrow \uparrow \Delta r \rightarrow \downarrow I \rightarrow \downarrow AD$.

3. **Deflation** is a general and consistent fall in the average price level. It is far more damaging to an economy than 'reasonable' inflation. There are two types of deflation:

   a. **Benign deflation** caused by increasing AS – there is deflation but rising GDPr

   b. **Malign deflation** caused by falling AD – prices and incomes both fall

4. **Counter-inflationary policies** are:

   a. **Contractionary** policies (e.g. policies aimed at decreasing AD) include *raising interest rates*, decreasing the supply of money, raising taxes on income and profits, *decreasing government spending*.

   b. **Supply-side policies** (increasing AS) include *lowering marginal income tax rates*, *easing up on labour market regulations*, *privatisation* of national industries, *decreasing union power*, *increased education* and skills in the work force... etc.

5. **Demand-side contractionary policies have negative effects** such as a decrease in government *tax receipts*, increased *unemployment*, *lower GDP* and personal income, time lags which make it difficult to time the policies correctly and the possibility of *creating recession* instead.

6. **Supply-side policies have some serious weaknesses**, such as actually increasing unemployment in the short run, can take a long time to implement and take effect, lower marginal tax rates can have serious consequences on the government budget, and studies show very limited effects of personal income tax cuts on labour supply.

MACROECONOMICS

54. Short and Long Run Phillips Curve

Key concepts: HL extension

- The SR Phillips curve – possible trade-off
- Shift of SR Phillips curve
- LR Phillips curve – no trade off in LR
- Natural rate of unemployment and full employment level of output

A study conducted by professor Phillips in 1958 showed a strong relationship between two of the most looked-at macroeconomic variables, unemployment and wage rates. The study, which covered the time period 1861 to 1957 in the UK, showed a clear *negative correlation* between the two variables. The relationship was to have great importance for further economic models after being slightly altered to show instead the relationship between **inflation and unemployment**.

and unemployment values for the UK over an almost 100 year period and based his line of best fit on the values. Wage increases were shown to have strong statistical correlation to decreases in unemployment. As wage increases ultimately feed through to the overall price level, a subsequent relationship between wage rates and inflation was found to hold steady. This provided the empirical and theoretical foundation for what became the Phillips Curve where higher inflation rates are associated with lower levels of unemployment.

The short run Phillips curve – possible trade-off

Figure 52.1 I shows Phillips's original curve, and Diagram II illustrates what ultimately became the graph we now refer to as the Phillips curve. Professor Phillips plotted all the wage
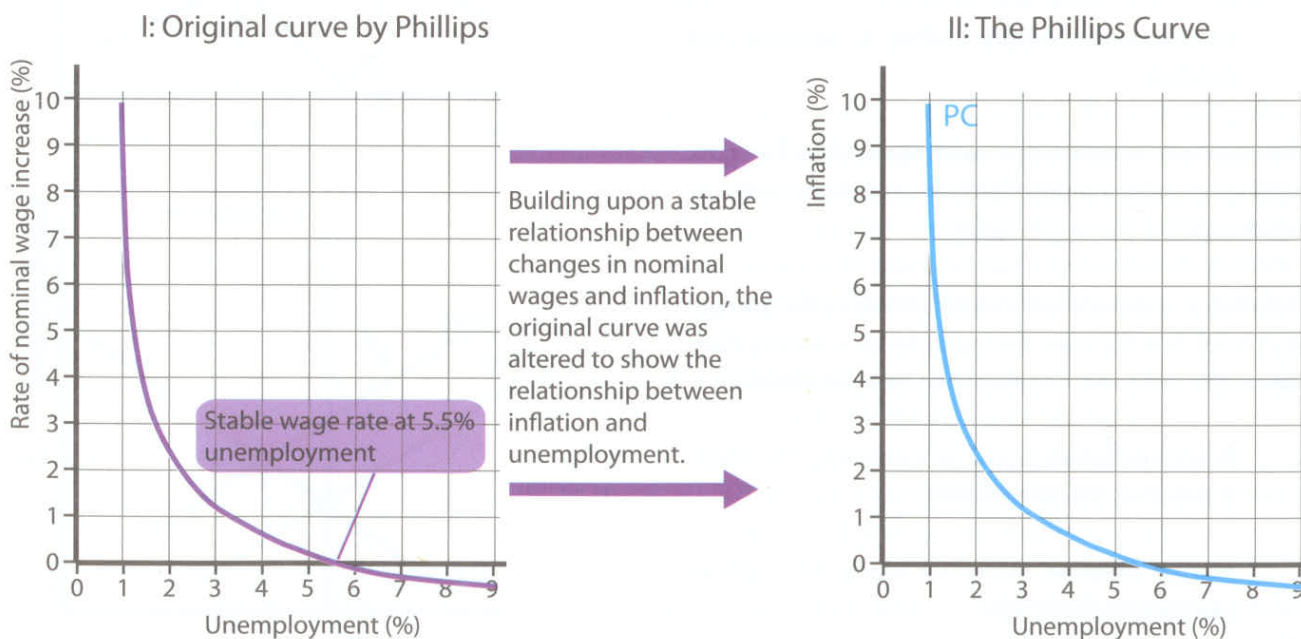

Figure 54.1 The original Phillips curve

The relationship between inflation and unemployment was subsequently found to hold broadly true in other countries during the 1950s and '60s. One rather remarkable fit of the Phillips Curve to actual values was the period between 1960 and 1969 in the US, shown in Figure 52.2.

Figure 54.2 Inflation and unemployment in the US, 1960 – 1969

(Source; *Abel & Bernake*, page 435)

Aggregate demand and the Phillips Curve

Having provided strong empirical evidence of correlation between inflation and unemployment, the results were incorporated into the economic theory of the time by a number of notable economists.[1] A main conclusion based on the trade-off was that while governments could apparently never fully conquer unemployment, there was at least an available choice for policy makers. The view that the Phillips curve showed a *stable relationship* between the price level (inflation) and unemployment had stark implications for economic policy. It provided governments with a 'menu' of possible policy goals shown by points **A**, **B** and **C** along the Phillips curve in **Diagram II**, **Figure 54.3**, where each level of unemployment/inflation would have its own specific opportunity cost. In simple terms, any given policy goal concerning aimed-for unemployment would have a trade-off in terms of inflation – and vice versa. Lowering unemployment from U_0 to U_1 (via, for example, expansionary fiscal policies – AD_0 to AD_1 in Diagram I) would entail the costs of higher inflation rates, i_0 to i_1. This is point **A** to **B** in both diagrams.
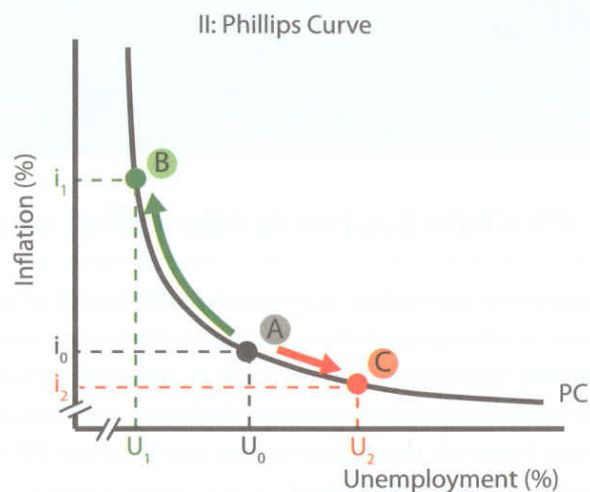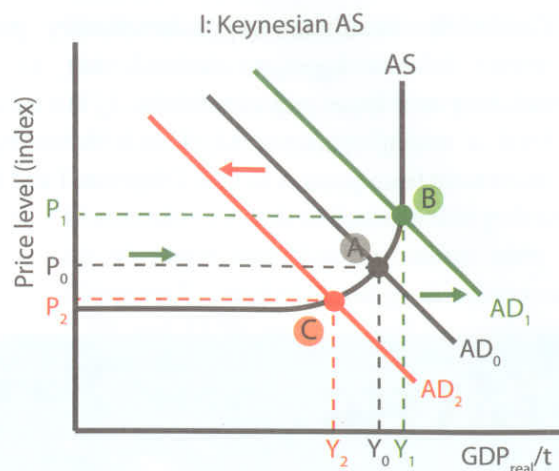

Figure 54.3 The AS curve and the Phillips curve

The evidence and conclusions rendered by the Phillips curve fit nicely into Keynesian theory since each point along the Phillips curve represented a macroeconomic equilibrium in line with the view of how the aggregate supply curve was constructed. Diagram I (Figure 54.3) shows how the 'inverted L' aggregate supply curve was adapted to show the trade-off. Each point along the Phillips curve in Diagram II has a corresponding movement along the aggregate supply curve.

- Assume that prices have been increasing at a stable rate at an output of Y_0 (Diagram I) and U_0 (Diagram II) whereupon *expansionary fiscal/monetary* policies are used to increase aggregate demand from AD_0 to AD_1. The increase in aggregate demand will increase national income from Y_0 to Y_1, and lower unemployment from U_0 to U_1 (Diagram II). This is shown in the diagram as a movement from point **A** to **B** on the Phillips curve, which shows that the cost of these expansionary policies is an *increase in inflation* from i_0 to i_1.

---

[1] Primarily *Paul Samuelson* and *Robert Solow* in 1960. Both became Nobel Laureates in Economics.

- Conversely, fiscal/monetary *contractionary policies* would decrease aggregate demand, AD_0 to AD_2, resulting in a lower rate of inflation, i_2, but a higher level of unemployment at U_2. This is shown by the movement from point **A** to **C** in **Diagrams** I and II.

Source: Mark A. Calabria at http://www.cato-at-liberty.org/is-there-an-inflation-unemployment-trade-off/ and the St. Louise Federal Reserve bank at www.stlouis.fed.org

A Little Depth

No Phillips Curve trade-off at all?!

For decades policy makers have pondered the costs of higher inflation in attempting to lower unemployment and this has also fuelled the Keynesian-New-classical debate. Skim back to figure 54.2 – surely this shows ample evidence of the trade-off? Well, yes. However, should one look at data since the 1960s the evidence is much less indicative.

Unemployment Rate (%)
Consumer Price Index
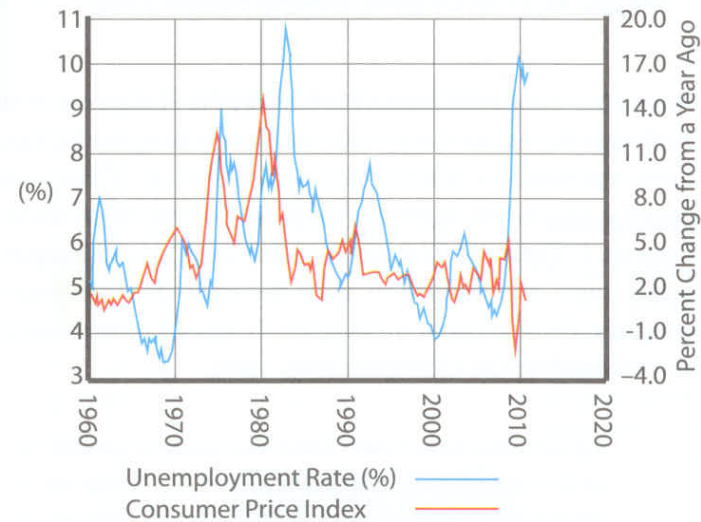
There is not much of a case for a Phillips-curve trade-off when one looks at the graph above. If anything, the correlation is often positive – there is ample evidence of inflation rising and falling roughly in concert with unemployment! Keeping in mind that correlation is not causality, the graph still doesn't lend a lot of support for policy makers being able to exploit a trade-off between inflation and unemployment. Back to the drawing board.

LR Phillips curve – no trade off in LR

During the 1950s and '60s governments – in line with demand management policies – used the Phillips curve trade-off as a premise in setting economic policies aimed at fine-tuning the economy. It was often easier to accept higher inflation rates than high unemployment, which contributed to rising inflation in many countries during the latter part of the 1960s. To make a simple point, the stable relationship between inflation and unemployment broke down in industrialised countries during the early 1970s. I once again illustrate this using inflation and unemployment figures for the US. Figure 54.4 illustrates how the stable relationship between inflation and unemployment in the US apparently ceased to exist.[2] (Compare with Figure 54.2.)

Figure 54.4 Inflation and unemployment in the US, 1970 – 1980

(Source: US dept of labour, Bureau of labor statistics at: ftp://ftp.bls.gov/pub/suppl/empsit. cpseea1.txt and http://stats.bls.gov/cpi/home. htm#data)

---

2 My defence in continuing to use US figures is that they are highly representative of what happened in most countries during the 1970s and '80s and that a great deal of what we now call mainstream theory has its origin amongst American economists and US data.

Notice, in Figure 54.43, how the most noticeable 'breakdown' occurs from 1973 onwards, where inflation doubled in 1974, unemployment almost doubled by 1975 – and both continued to spiral outwards for the rest of the decade. The primary cause, you have no doubt already guessed, is to be found in the first oil crisis in 1973/'74 and the next one in 1979/'80. Diagrams I-III in Figure 54.5 are based on the values given in Figure 54.43 and serve as an illustration of what happened.

I: Initial supply-side shock and stagflation in 1973–74 …

II: … and cost-push spiral during 1970s

III: Phillips curve illustration of stagflationary period

Figure 54.5 I – III; Supply-side shock, cost-push inflation and stagflation in the US, 1970 – 1980
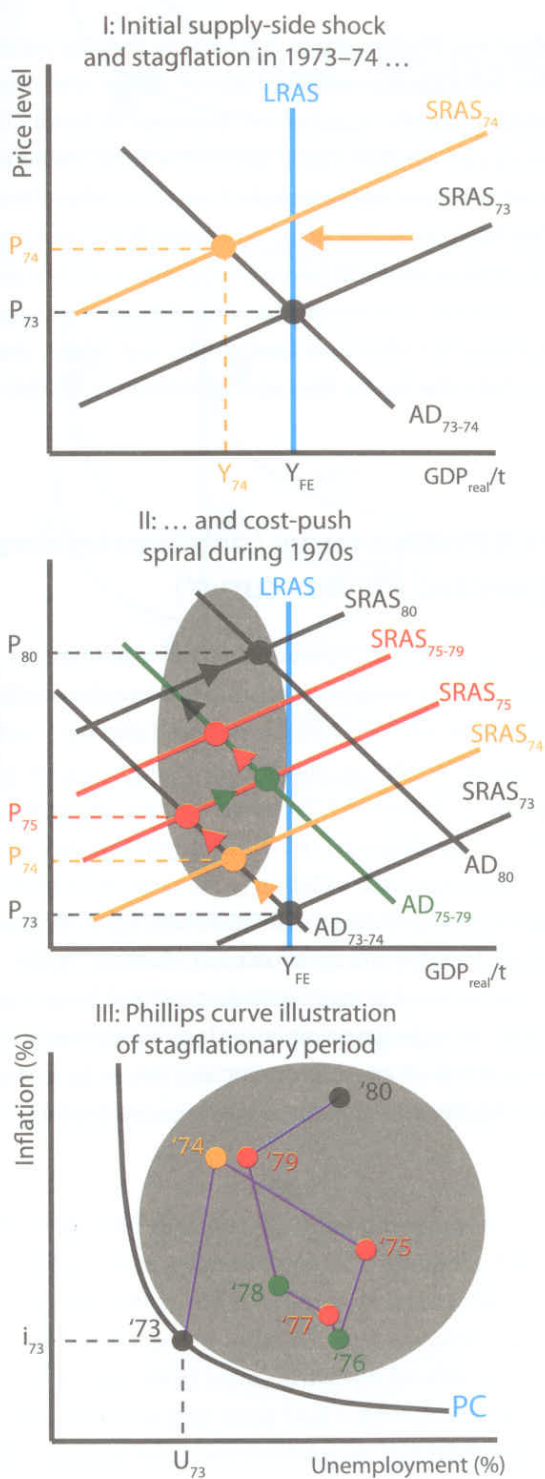
- The initial supply-side shock in 1973/'74 (Diagram I) caused inflation to rise from 6.2% in 1973 to 11% in 1974, while unemployment increased from 4.9% to 5.6% (and to 8.5% in 1975).

- Consecutive periods of demand-side policies, bidding up of wages and thus increasing costs for firms (Diagram II) resulted in cost-push inflation of between 5.5% and 13.5% annually, and unemployment levels hovering between 6% and 9%.

- Diagram III illustrates this process in an 'outward bound spiral' in terms of the Phillips curve, where it is quite evident that the relationship between inflation and unemployment has collapsed, resulting in stagflation.

The breakdown of the relationship depicted in the Phillips curve is a rare example of when economic theory precedes real events. During the latter part of the 1960s, two economists, **Milton Friedman** and **Edmund Phelps**, came to the conclusion that the *Phillips curve trade-off could only work in the short run* and thus that the stable relationship depicted could not last. The Friedman-Phelps proposition form the '60s turned from theory-based prediction to empirically-based 'truism' during the stagflationary period of the 1970s and early '80s.

MACROECONOMICS

The new-classical criticism of the short run Phillips curve

Assume an economy in equilibrium at Y_0 in **Diagram I** (**Figure 54.6**) where the natural rate of unemployment is U_0 and yearly inflation is i_0, **Diagram II**. Aggregate demand increases from AD_0 to AD_1, and the economy moves from Y_0 to Y_1. In moving from Y_0 to Y_1, inflation is actually eroding real wages, since the rate of inflation has increased from i_0 to i_1, shown in **Diagram II**. Say that labourers have negotiated for wage increases of 3% for the coming time period but that inflation during the time period turned out to be 3.5%. This means a real wage loss of 0.5%. If workers are unaware of this fact, then they suffer from **money illusion** (i.e. that the nominal wage increase is real) and will ultimately adjust demand and spending to real wages. Aggregate demand will fall back to AD_0 and deflationary pressure will lower inflation and increase unemployment. Thus the economy would move from point **B** (in **Diagram II**) along the short run Phillips curve back to i_0 and U_0 at point **A**.

I: Expansionary increase in AD
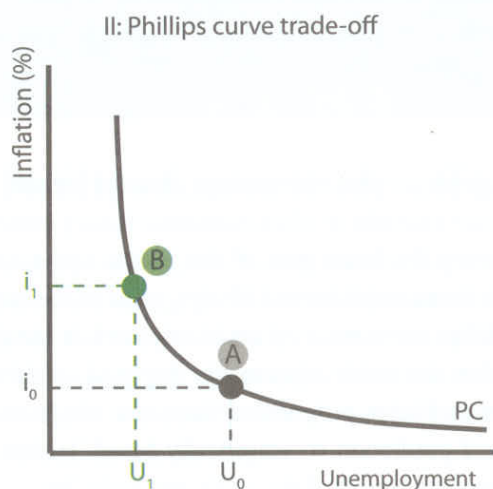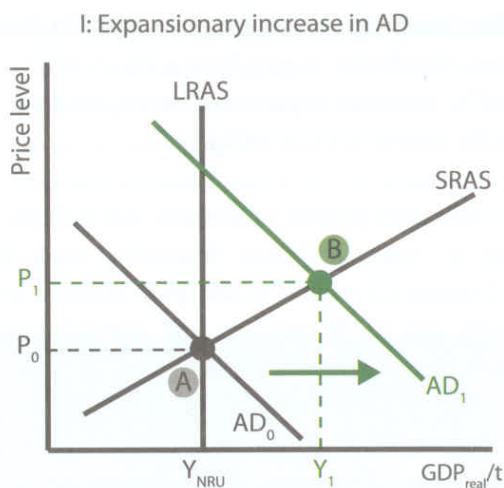

II: Phillips curve trade-off

Figure 54.6 What Friedman/Phelps reacted to

Labourers do not suffer from 'money illusion'

It was precisely the above scenario that Friedman and Phelps opposed. According to their new-classical/ monetarist view, labourers do not suffer from money illusion and will therefore *adapt* to changing inflation rates by altering their **inflationary expectations**.[3] As pointed out in Chapter 46 (*The new- classical view of long run aggregate supply*), economic agents (producers and consumers) act strongly upon expectations.

A silly parallel: Getting your breakfast together in the morning, you will open cupboards and drawers where you expect to find things and your actions are based on expectations which in turn have been strengthened by past experiences; the bread is *always*

in the right-hand cupboard. If your mom/maid/spouse one day changes where everything is kept you will quickly adapt ... depending on how quickly your brain wakes up in the morning. "Ah, it's now kept in the bin by the stove!" Next morning you go for the bread bin. You will act similarly in your wage demands if you expect inflation rates to increase – e.g. you will demand higher wages if you expect your real wages to have deteriorated by an increase in inflation.

The short run Phillips curve is similar in that the curve is based on the inflationary expectations of firms and households. Expected inflation is based on previous inflation rates, and the short run Phillips curve governs a time frame where the expected inflation rate does not change, i.e. where the inflation rate has been such that firms and households assume that the current time period will have a predictable and thus *anticipated* rate of inflation. This means that when labourers engage in wage negotiations in order to at least retain real wages, the premise will be that inflation for the coming period is a known constant.

The LR Phillips curve (or 'expectations-augmented Phillips curve')

Assume an economy operating at full employment and at a natural rate of unemployment of 5% and where inflation has held steady at 3% for a longish period. Labourers' inflationary expectations render a short run Phillips curve of $SRPC_1$ (**Diagram I, Figure 54.7**) based on an expected inflation rate (i^{exp} in the diagram) of 3%. If inflation is indeed 3% then anticipated (expected) inflation is the same as de facto inflation and labourers and firms will have correctly increased both output and factor prices to market clearing levels. In other words, an increase in aggregate demand has been matched by a decrease in aggregate supply so that the economy remains at the natural rate of unemployment and prices have risen by the anticipated 3%. This is point **A** in **Diagrams I** and **II**.

---

3   This is why the theory is referred to as 'expectations augmented'; augmented means 'enhanced' or 'supplemented'. In other words, economic agents' behaviour is **augmented by inflationary expectations**.
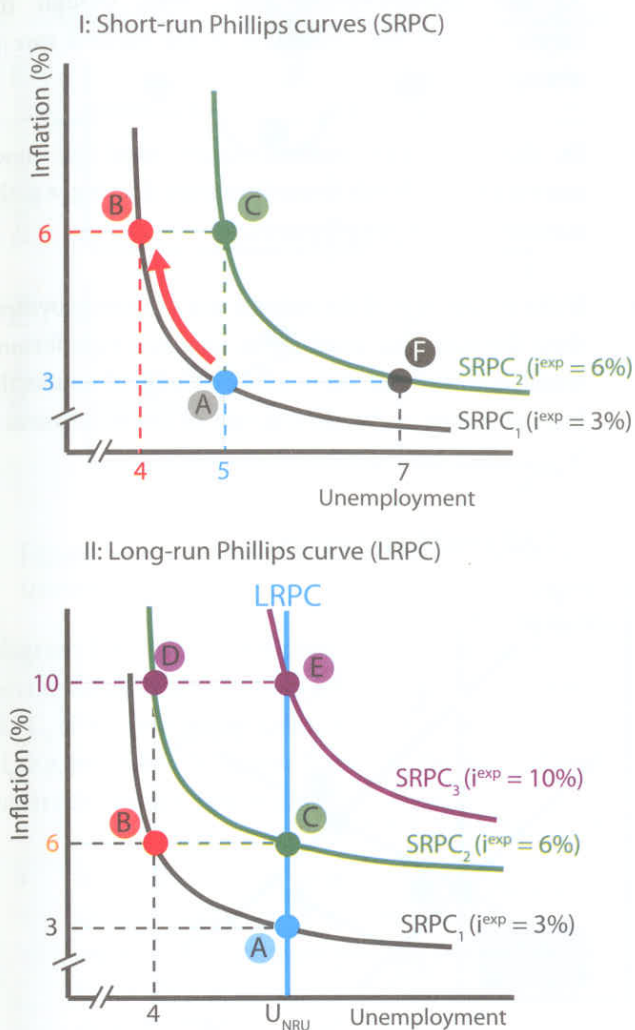
I: Short-run Phillips curves (SRPC)

II: Long-run Phillips curve (LRPC)

Figure 54.7 Short and long run Phillips curve

- Now, government has achieved its aim of lowering unemployment to 4%, but since *labourers do not suffer from money illusion*, point **B** is not a long run equilibrium point. In due course labourers and unions **adjust inflationary expectations to 6%** and will bid up wages to compensate for loss of real income. This will result in higher production costs and dissolve the previous real increase in profits for firms.

- **B to C:** As real costs become apparent to firms, they will cut output and *decrease their demand for labour*. Simultaneously, labourers who initially offered themselves on the labour market under the illusion that real wages were higher than turned out will withdraw from the labour market. This will cause both markets – goods and labour – to clear, leading to point **C** where the economy is once again in equilibrium and unemployment has increased to the *natural rate of unemployment*. Economic agents now operate under inflationary expectations of 6%, i.e. along a new short run Phillips curve, **SRPC₂ (i^{exp} = 6%)**.

If no further stimulatory action is taken the economy will remain at full output and the natural rate of unemployment at point **C**, now shown in **Diagram II, Figure 54.7**, at a 6% rate of inflation. **Diagram II** shows the long run Phillips curve which is based on a number of short run curves at specific rates of *expected* inflation. At each new level of inflationary expectations, equilibrium is restored when unemployment matches the natural rate of 5%, i.e. points **A, C** and **E** along the long run Phillips curve **LRPC**. If workers continuously adjust inflationary expectations upward in accordance with stimulatory policies, the short run Phillips curve will move ever outward/upward. The only way for demand side measures to once again reduce unemployment is for government to 'fool' workers again – e.g. by using even more fiscal/monetary stimulation.

The **long run Phillips curve** indicates that in order to keep unemployment *below* the natural rate, ever higher inflation rates will be necessary in order to keep a degree of money illusion present by continuous monetary expansion and/or fiscal stimulus. This is indicated in **Diagram II** as the dotted purple arrow leading upwards from point **B** with 6% inflation to point **D** with 10% inflation.

What if **deflationary monetary policies** are implemented in order to deal with inflation? Assume that at point **C** in **Diagram I, Figure 54.7** contractionary monetary policies are implemented, thereby reducing aggregate demand. Assume the

Posit instead that firms/labourers anticipate 3% inflation but that government considers a doubling of inflation (to 6%) a reasonable cost of lowering unemployment by one percentage point. By **stimulating aggregate demand** (say by lowering interest rates), firms will increase output to meet the increase in demand and start to hire more workers. Referring to **Figure 54.7**:

- **A to B:** The decrease in unemployment moves the economy from point **A** to point **B** along the **SRPC₁** in **Diagram I (Figure 54.7)**. Since de facto inflation now exceeds anticipated inflation, and workers' wages are fixed for the short term, firms will be able to increase their profit margins since final output prices have risen but nominal wages are unchanged. Note that wage rates and wage increases were set when both employers and labourers anticipated 3% inflation!

tight monetary policy is effective and decreases de facto inflation to 3%. This means that households/firms have overestimated inflation. Firms' real costs rise since wages levels were based on the premise of 6% inflation and aggregate supply will slow more than aggregate demand, leading to layoffs and increased unemployment. The economy is at point **F** in **Diagram I** and assuming that the inflation rate remains constant over the next time period, households and firms will ultimately adapt inflationary expectations to the actual rate of 3%. Labourers will lower wage demands enabling the labour market to clear and the economy returns to equilibrium at point A. While this may sound relatively straightforward, keep in mind that the 'journey' from point **C** to point **A** entails passing by point **F**, marked by the considerable social costs of two additional percentage points of unemployment for an indeterminate time period.

Natural rate of unemployment and full employment level of output

The conclusion of the expectations augmented Phillips curve is that there is **no trade-off in the long run** between inflation and unemployment. Attempts to reduce the rate of unemployment by demand side measures will fail in the long run since each successive stimulatory package aimed at *permanently* reducing unemployment would ultimately lead the economy back to its natural state but at a higher price level. There would be a separate Phillips curve for each anticipated level of inflation, rendering a **vertical long run Phillips curve** most similar to the long run aggregate supply curve. In fact, strictly speaking, they portray the same issue; any stimulatory measure to the economy which does not have a foundation in real (potential) output will not increase output or lower unemployment in the *long* run.

The congruence (= correspondence) between the AS-AD model and the long run Phillips curve is shown in Figure 54.8, diagrams I – IV, in an attempt to put the pieces together for you while illustrating the natural rate of unemployment.[4] In following the iteration below, keep in mind the following points:

- When the labour market is in equilibrium then only structural, frictional and seasonal unemployment exist.

- As this unemployment exists even though the labour market has cleared, it is the *natural rate of unemployment.*

- In the long run market forces tend to move unemployment levels towards market clearing, e.g. the natural rate of unemployment.

- If the economy is at the natural rate of unemployment then the economy must be in long run equilibrium, which is to say that AD = LRAS. This is of course the level of output at the natural rate of unemployment – Y_{NRU}.

I: AD and AS in long run

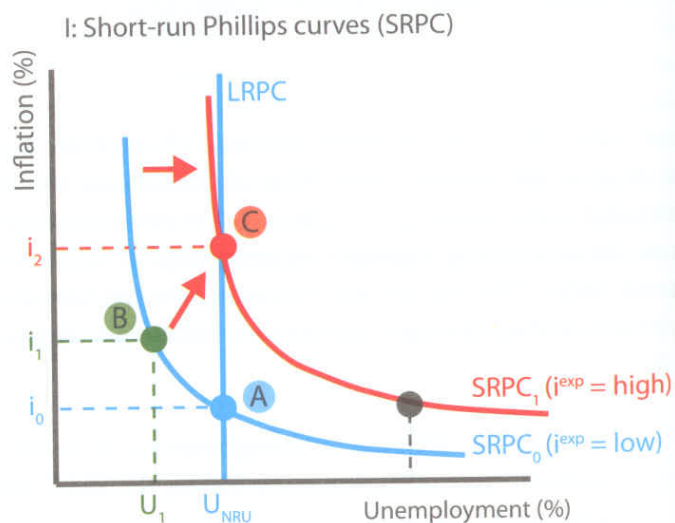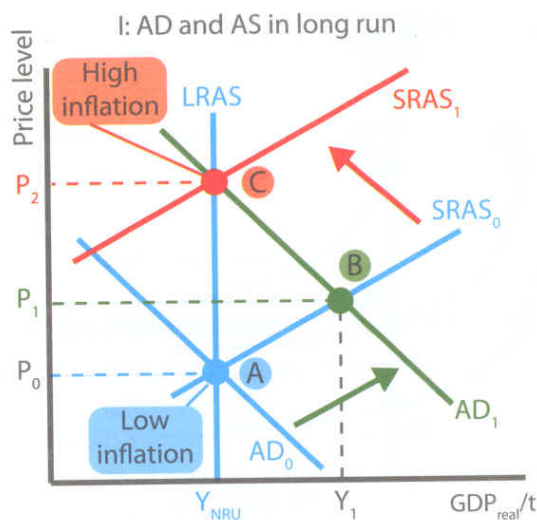
I: Short-run Phillips curves (SRPC)

Figure 54.8 I & II AS-AD model and the natural rate of unemployment

---

4 I use these diagrams with grateful acknowledgement to Professor Klas Fregert and associate professor Lars Jonung at Lund University.

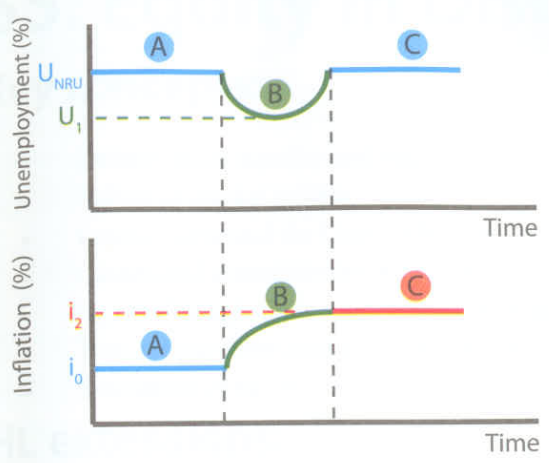III & IV: Time series for unemployment and inflation

Figure 54.8 III AS-AD model and the natural rate of unemployment

Diagram I in **Figure 54.8** shows how an economy is initially moving along at a rate where inflation is correctly anticipated, $SRPC_0$ (i^{exp} = low) in **Diagram II**. The resulting rate of inflation is i_0, e.g. point A in **Diagrams II** and **IV**. This *general equilibrium* renders the natural rate of unemployment.

- **A to B**: If government now stimulates aggregate demand (AD_0 to AD_1, **Diagram I**) there will be *unanticipated* inflation resulting from the demand side shock. This is shown by a movement along the initial short run Phillips curve $SRPC_0$ (in accordance with earlier arguments) and the economy moves from point A to B.

  o The change is illustrated further in **Diagram III** (unemployment falls to U_1) and **Diagram IV** (increasing rate of inflation).

- **B to C**: As labourers do not suffer from money illusion, they bid up wages in adjustment to higher inflationary expectations. The increase in labour costs for firms causes aggregate supply to decrease to $SRAS_1$ and point C in **Diagram I**. Jobs are lost while the economy experiences yet another increase in the rate of inflation before returning to the natural rate of unemployment at Y_{NRU}.

  o **Diagram II** shows how a new short run Phillips curve, $SRPC_0$ (i^{exp} = high), is formed at an anticipated inflation rate of i_2. **Diagram III** shows how unemployment has returned to U_{NRU} at point C. **Diagram IV** illustrates how inflation now progresses at a higher rate, i_2.

The progression from point A to B to C in Figure 54.8 explains the new-classical view of how the economy will always tend towards the natural rate of unemployment in the long run. The vertical long run Phillips curve tells policy makers that there is no long run trade-off between inflation and unemployment. If the aim is to decrease unemployment below the natural rate of unemployment in the *long run*, then the appropriate policy action is supply-side policy.

POP QUIZ 54.1

The CPI and the Phillips curve

1. Why are goods in the CPI weighted?

2. What are the main weaknesses of using the CPI as a measure of inflation?

3. How does 'money illusion' help to explain the trade-off shown by the Phillips curve?

4. Illustrate how increased inflationary expectations would lead to higher inflation.

5. Using the long run Phillips curve, explain how demand side policies would be unable to decrease the natural rate of unemployment.

6. How might supply-side policies affect the natural rate of unemployment?

369

Summary & revision

HIGER LEVEL

MACROECONOMICS

1. The **short run Phillips curve** (SRPC) indicates a *trade-off between inflation and unemployment* – falling unemployment has been seen to correlate to rising inflation.

2. The short run Phillips curve (SRPC) was incorporated into Keynesian thinking in the 1950s to show the link to AD and a possible '**menu' of different inflation-unemployment options**.

3. **Freidman and Phelps**, working in the 1960s, criticised the SRPC, claiming that people did not suffer from money illusion and that *increases in AD beyond the natural rate of unemployment would cause inflationary bidding up of wages and a decrease in SRAS*. The economy would ultimately revert to the level of income at the natural rate of unemployment.

4. The supply-side shock and resulting **stagflation** (stagnant economy + inflation) in the mid-1970s meant a break-down of the short run Phillips curve and seemed to corroborate the existence of a vertical long run Phillips curve.

5. The **LRPC** is based on the concept of **adaptive expectations**; people adapt to inflationary expectations. An increase in AD beyond the natural rate of unemployment causes inflation and a decrease in real wages. It also fuels households' inflationary expectations. The resulting bidding up of wages as households and labourers adapt, shifts the SRPC to the right and ultimately general equilibrium is restored along the LRPC. Unemployment has returned to the natural rate of unemployment but inflation has increased to a higher level.

6. The possibility of a LRPC has some weighty **ramifications for economic policy**. The supply-side school of thought has the concept of LRAS and a natural rate of unemployment as a cornerstone for positing that only increases in long run aggregate supply can increase real income in the long run.

55. Equity in Distribution of Income

Key concepts:

- Definitions of equality and equity
- Failure of market system
- Lorenz curve and the Gini coefficient
- Causes and consequences of poverty
- Equity through taxes – redistribution of income
- Transfer payments and services in kind
- Evaluation of equity policies

HL extensions:

- Calculating the marginal tax rate
- Calculating the average tax rate

"The avoidance of taxes is the only intellectual pursuit that still carries any reward." J. M. Keynes

Definitions of equity and equality

One of my favourite baiting-games I play with younger students is to claim that there is nothing wrong with a society where people starve alongside the extremely rich. I sometimes refer to the gap in wealth and income to be found in, say, Brazil, and claim that this is perfectly acceptable in terms of economics. We have, after all, solved the basic economic problem of who gets what – we allow income and wealth to decide. I do this towards the end of class in order to make my escape and avoid the inevitable screaming and general outrage. I make sure that the whiteboard is full so that they are preoccupied as I sneak out. When the students have cooled down a bit, usually after a few days, I carefully distinguish between 'equity' and 'equality'. Equity in economics is a concept meaning 'fairness' and 'justice', for example, that everybody should have the same right to work, own property and start a company – regardless of gender, ethnicity etc. There would then be fairness in the distribution of output, e.g. a just proportion of wealth to each and every citizen. Equality, on the other hand, would mean that everyone would have the same ability to work, own property and start a company – all would get equal portions of the wealth created.

Definition: Equity'
The concept of **equity** in economics deals with the highly normative concept of fairness in the distribution of wealth and income. Most countries have customs, laws and traditions aimed at giving disadvantaged members of society 'fair shares'.

Definition: 'Equality'
Equality deals with spreading wealth and income equally, regardless of position or income in society.

The concept of equity is naturally highly normative and will therefore vary greatly over time and between cultures. There are different moral and ethical bases in societies. In line with these differences, governments attempt to create equity, which is to say 'everybody should have the same chance of bettering his or her situation'. Economics must often deal with the questions of equity:

- Should the rich be taxed proportionately more than the poor and middle-classed? Should they pay more for university education?

- Is it better to tax consumption (value-added taxes, VAT, for example) or income (taxes on profits and dividends)?

- To what extent should the socially/economically disadvantaged be given additional resources at the expense of the economically advantaged – i.e. transfers of income and wealth aimed at evening-out income disparities?

- What about the viewpoint of the wealthy; is it not an economic fact that the goods and services we desire are better furthered by self interest and the profit motives of firms? Is it more economically efficient to *increase* output rather than *re-distribute* it? Should not governments try to create a favourable environment for firms and a framework for enterprise in order to increase the cake from which all subsequent taxes and income transfers will be sliced? Wouldn't society ultimately gain if the incentives to produce (say via low corporate taxes) were enhanced?

Failure of market system

The above examples are but a small portion of a recurring debate within both economics and politics, dealing with the desirability and efficiency of government intervention in matters of equity. We have returned to the issues outlined in production possibility curves earlier, but now with a slightly more ideological and thus political flavour. The appealing simplicity of 'invest now in order to produce more in the future' must be appended by realism: it is not at all certain that any part of 'more' will be distributed equitably or evenly. Our basic economic problem is not so basic after all. If output were to grow by 50% and go to 0.1% of the population then we would certainly face some sort of sharp societal reaction – from protest lists to burning tyres in front of parliament. Most countries will have a redistribution system built into government policy, which evens out income differentials to a certain extent. Governments can redistribute income in three basic ways; *taxation*, *transfer payments* and *goods* and *services in kind*. (See further on.)

Lorenz curve and the Gini coefficient

"Wal-Mart ... do they like, make walls there?"
Paris Hilton

One of the many controversial areas in economics is the issue of income inequality *within* societies (often over a time period) and *between* societies. Media coverage commonly puts this in terms of a 'gap between rich and poor', or 'inequality in country X'. For example, an article in the *International Herald Tribune* states that in 2000 the richest 1% of Americans had more after-

tax income than the bottom 40%, representing 15.5% of total national income – triple the level that this group had 20 years earlier.[1] Another article, in *Time Magazine*, points out that in 2003 the wealthiest 1% of Americans accounted for 40% of total wealth (note, not income) while the corresponding value in the UK was 18%.[2] Figure 55.1 below shows a simple and graspable way to compare inequality within and between countries, namely the Lorenz curve.

Lorenz curve

In Diagram I (Figure 55.1), the Y-axis shows the cumulative (= collective, summed-up) percentage of total income and the X-axis shows the cumulative percentage of all wage earners. The 45 degree line is the line of perfect equality, i.e. a country where 1% of income goes to 1% of wage earners, 2% of income goes to 2% of wage earners and so forth along the line. The upward-sloping curve is a Lorenz curve, in this case showing that income distribution is rather uneven. The farther away the Lorenz curve is from the line of perfect equality, the more unequal the income distribution. In the example above, the bottom 20% (the first quintile) of wage earners accounts for just 2.8% of total income; the second quintile accounts for 6.4% of income (9.2% - 2.8%); and skipping the next two, the top quintile of wage earners accounts for 61.1% of all income.

I: Lorenz curve
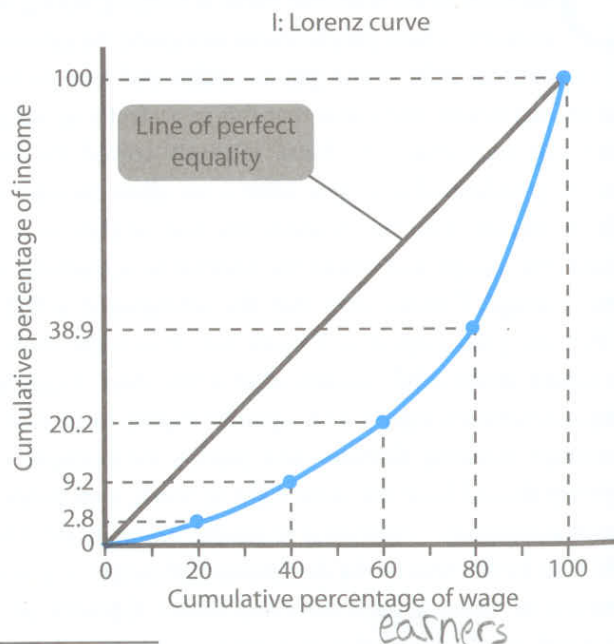
1   IHT, US rich get richer, and poor poorer, data shows, 25 Sept 2003 (During 2005 the top 1 percent of USAs richest people earned 21.2% of total income. Wall Street Journal in Oct 2007, quoted in http://www.privataaffarer.se/newsText. asp?src=pa&a=23038)

2   *Time Magazine*, 17 Nov 2003, page 22 (Official US government figures at the US Census Bureau show clearly that income inequality has been increasing from 1947 to 2007. See http://www.census.gov/hhes/www/income/histinc/f04.html)
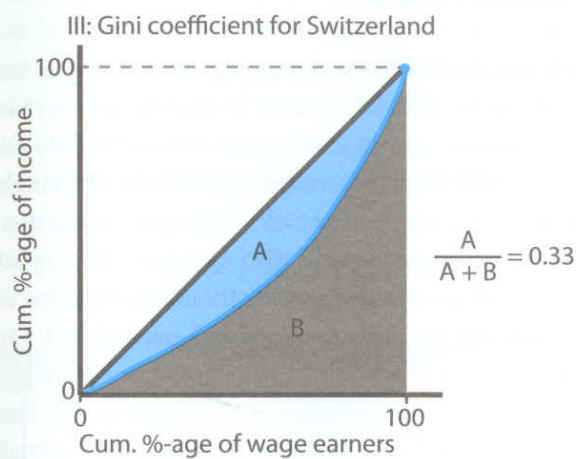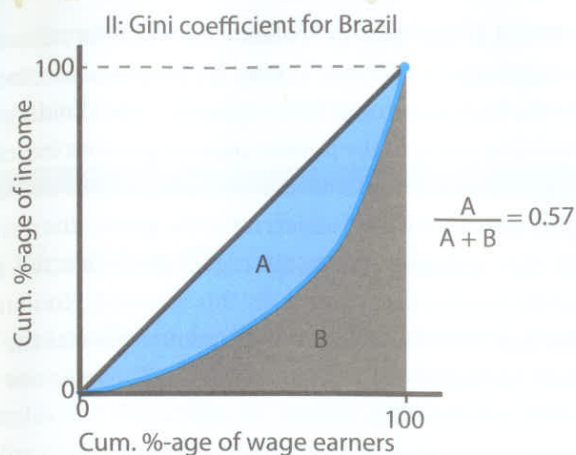
II: Gini coefficient for Brazil

$$\frac{A}{A+B} = 0.57$$

III: Gini coefficient for Switzerland

$$\frac{A}{A+B} = 0.33$$

Figure 55.1 I – III Lorenz curves

Gini coefficient

How far removed from reality is this example? Well, actually, I have just described Brazil, notoriously one of the most unequal societies in the world. Now, if we wish to compare income distribution in Brazil with other countries (or indeed Brazilian income distribution over time), we can calculate the so-called **Gini coefficient**, which is a value of the 'distance' of the curve from the 45 degree perfect equality line. This is calculated by taking the ratio between area A and A+B (Diagrams II and III in Figure 55.1).

If income were perfectly distributed then the Gini coefficient would be 0. If one wage earner accounted for 100% of income, the coefficient would be 1. The *higher* the Gini coefficient, the more *unequally* income is distributed. According to the World Bank in 2007, Brazil had a Gini coefficient of 0.57 while Switzerland had 0.33. The table in Figure 55.2 shows these and a few other Gini coefficients for comparison.

Country [year of survey]	Gini coefficient	Percentage share of income		Kuznets ratio [(2) / (1)]
		Lowest quintile (1)	Highest quintile (2)	
Brazil [2004](see diagram)	0.57	2.8%	61.1%	21,8
Honduras [2003]	0.56	3.4%	58.3%	17.1
Jamaica [2004]	0.38	5.3%	51.6%	9.7
Sweden [2000]	0.25	9.1%	36.6%	4.0
Switzerland [2000] (see diagram)	0.33	7.6%	41.3%	5.4
Luxembourg [2002]	0.27	9.4%	36.5%	3,9

(Sources: World Development indicators 2007, at World Bank, at http://siteresources.worldbank.org)

Figure 55.2 Gini coefficients, selected countries

The Gini coefficient is a most convenient way of summarising the degree of income inequality in a country, lending readily available figures for comparison. It is immediately apparent that Brazil is indeed a most unequal country in terms of income distribution, rendering a Gini coefficient of 0.57. I have included three of the richest countries in the world as a contrast, where it should come as no surprise that Sweden's fair-minded egalitarian philosophy and high tax rates result in one of the lowest Gini coefficients in the world; 0.25. The country which probably has the most uneven distribution of income in the world is Namibia, with an estimated Gini coefficient of 0.74. Data from a population census taken in 1994 pointed to the richest 7,000 people (out of a population of 1.4 million) having the total income of the poorest 800,000.

At the other end of the scale, the country with the lowest Gini coefficient in the world seems to be Denmark at 0.247. Keep in mind that the Gini coefficient only measures the *relative* distribution of income and not poverty levels. There are also a few **weaknesses with the Gini coefficient**:

- It is quite possible that two countries with entirely different income distributions have the same Gini value.

- Also, the shape of the curve gives information only on total distribution and not on the relationship between the richest and poorest segments (see Kuznets ratio below).

- Finally, a country with a high Gini coefficient might in fact have a generally high standard of living in low income groups – just as it is equally possible for a country with a low Gini coefficient to have widespread poverty across most income groups.

OUTSIDE THE BOX

The Kuznets ratio and Kuznets curve

The Kuznets ratio

I have included three additional Outside the Box columns in the table in Figure 55.2 to show the Kuznets ratio, originating with economist Simon Kuznets. This ratio of inequality is a

measurement of the 'span' or 'distance' between the richest and poorest segments in a society. Taking the proportion of income going to the highest earning 20% - quintile – and dividing it by the proportion going to the poorest quintile gives an indication of the gap between the richest and poorest portions of society. The higher the value of the Kuznets ratio, the greater the distance between rich and poor – e.g. the higher the value the more unequal the income distribution. By this measure, Honduras is more unequal than Brazil in terms of income whereas the Gini coefficient indicated the reverse. Sweden still comes out near the top with a ratio of 3.6.

The Kuznets curve

Another – highly controversial and thus infinitely more famous – concept put forward by Kuznets is that there is correlation between economic growth and income distribution. He posited that during initial stages of economic growth the distribution of income would worsen, i.e. the rich would gain proportionately more of the increased income than the poor. During later – unspecified – stages of growth the distribution of income would improve, i.e. the gap between rich and poor would narrow. Plotting income against income distribution (measured by the Gini coefficient) gives an 'inverted U', known as the Kuznets curve, where growth in per capita income results initially in worsening income distribution, i.e. a higher Gini coefficient. At some – unspecified – level of income, correlation will reverse and income distributions will improve.

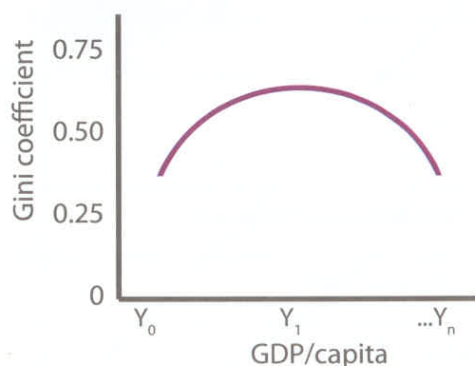Figure 55.3 Kuznets curve

The Kuznets curve has been fiercely debated for over 40 years now. The proposition that income distribution would at first widen and then narrow was widely studied and by the 1970s had broad acceptance. Yet a good many later empirical studies show little evidence of either positive or negative systematic relationships between per capita growth and income distribution.

The controversy continues unabated around a curve which Kuznets himself claims was " … 5% empirical … 95% speculation"!

The World in a Champagne glass

The Gini coefficient for the world has remained constant at 0.67 since the 1970s. The Human Development Report of 2005 illustrates the disparities in income between the rich world and the poor world by computing the richest 20% against the poorest 20% of the world's population. The average income of the top 20% is roughly 50 times higher than the average of the poorest 20%. The 'Champagne glass' illustration (Figure 5.2.7) shows that the top 20% of account for 75% of world income. The bottom 40% – the 'stem' of the glass – account for 5% of world income. The bottom 20% account for 1.5%.[3]
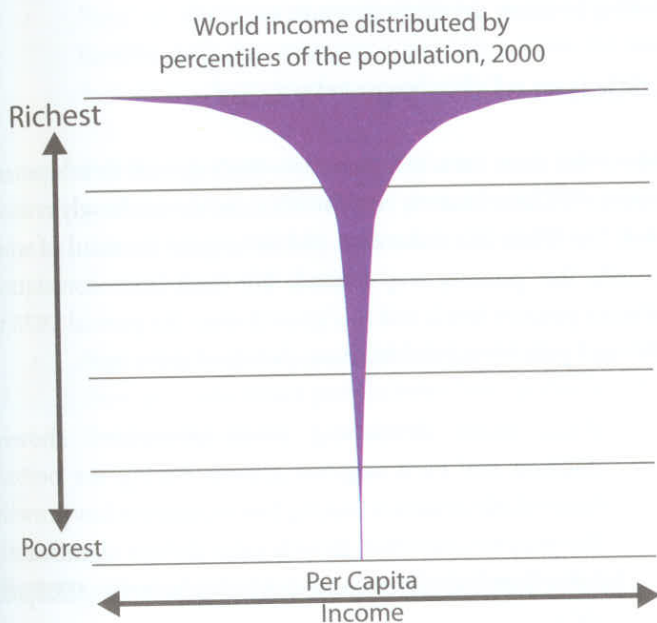

World income distributed by percentiles of the population, 2000

Figure 55.4 World income distribution, 2000 (Source: HDR 2005, page 37)

Causes and consequences of poverty

The term 'poverty' is one we all know … or *think* we know until asked to define it. On a general level, poverty is a situation where people lack basic necessities such as food, clothing and shelter. We would refer to this as **absolute poverty**, commonly defined as a level of income or consumption falling below the minimum level required to meet basic needs, creating a 'poverty line' allowing between-country comparison.

3 HDR 2005, page 36.

In an attempt to at least have some standardised bar of poverty whereby different countries could be compared more objectively using a common unit, the World Bank has set *international poverty lines* at $US1 and $US2 per day in purchasing power parity terms. Using the World Bank's updated '1.25 $US a day' as a lower limit, the World Bank estimates that 1.29 billion people (22.4% of the population in LDCs) in the world could be defined as living in absolute poverty in 2008.[4] Setting the bar at $US2 the number increases to 2.8 billion people. The problem in using this absolute measure is that what is considered 'minimum level required' will vary over time and in different societies, so even the absolute poverty line will change – something the World Bank carefully points out.[5]

Relative poverty is the result of uneven income distribution, where the lowest income earners are compared to other groups. This definition is of course highly normative and will vary greatly between different countries, depending on income levels and also upon how 'poverty' is defined. The UK defines relative poverty as having an income below 60% of the median net disposable income (income after tax plus transfer payments) while the US sets the bar at 50%. 'Relative' in its definitional context of poverty means that the poverty bar will be set according to domestic norms. Thus, a poor person in Iceland – shown in recurring surveys to have one of the highest perceived quality of life in the world – would have a king's ransom as a monthly income according to a squatter on the outskirts of Mexico City. This makes it essentially impossible – and pointless – to compare different countries in terms of poverty levels.

> ### Definition: 'Absolute poverty'
> When people lack the basic resources required to meet basic needs such as food and clothing, one speaks of **absolute poverty**. The World Bank uses a basic metric for a cross-country definition; 1.25USD per day.
>
> ### Definition: 'Relative poverty'
> **Relative poverty** means having an income below a set median or average, say less than 50% of the median income. This normative concept varies between countries.

4 See: http://www.worldbank.org/poverty/mission/up2.htm
5 IBID (See also the full text of the 2001 WDR on Poverty at http://www.worldbank.org/poverty/wdrpoverty/report/index.htm

Perhaps the most obvious indicator of lack of development is **poverty**. The UN Millennium Goals of 2000[6] set down a number of development goals, and at the top of the list was the goal to reduce the amount of people living on less than $US1 (at PPP) per day by 50% by the year 2015. There were close to 1.2 billion people living below this absolute poverty line at the time of the Millennium Summit, and many of them face a predicament known as the **poverty cycle** or **poverty trap**.[7]

Poor people will have a notably low propensity to save, as there will be little enough to spare after basic necessities have been purchased. In other words, savings will have a very high opportunity cost in the form of foregone – vital – consumption. Recall that investment is the mirror image of saving (Section 3.1) and that banks/financial institutions facilitate this flow of funds from households to firms. Economic theory posits that households' willingness to put aside present consumption in order to increase future consumption is based on income levels. The cycle of poverty is thus:

Low savings in developing countries results in…

- … scarcity of investment funds – the investment funds needed by firms to increase output and build infrastructure – and **low investment,** …

- … which is central to a country's output potential, will hamper economic growth, e.g. result in **low national income** …

- … and since income provides the proportion used for savings, there will be **low** …
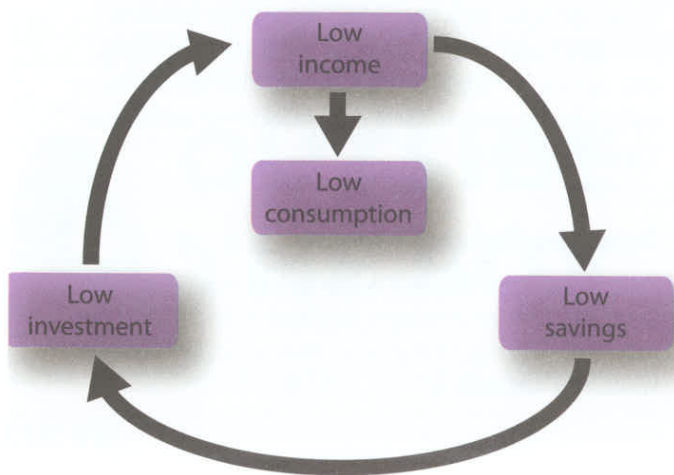

Figure 55.5 The poverty cycle

The cycle is enhanced by two additional forces:

1. Firms' investment plans are to a certain extent based on predicted **consumption** levels by households and investment by other firms. A low level of consumption will feed through to continued low investment levels.

2. Low investment (in both fixed and human capital) will stifle **productivity** gains in the economy and keep real wages low. This too will exert a negative force on incomes and consumption.

Figure 55.5 looks like a merry-go-round from hell or the Hotel California; you can jump on anywhere but you can't get off. Yet there are ways for developing countries to get off the negative cycle, for example by creating sound financial institutions and attracting foreign investment. Unfortunately, these and other possible solutions require working governments, political stability, and the rule of law – some or all of which are noticeably lacking in many developing countries.

Criticism of the 'poverty trap'

One of the most critical economists dealing with development issues is William Easterly (see for example his scathingly critical book *The White Man's Burden*) and he is quite scornful of what he calls the 'poverty trap legend'. He finds little correlation between poverty levels and low growth over the period 1975 to 2001 and puts forward three main points of criticism[8]:

1. While many developing countries indeed showed stagnant and even negative growth during the period, there is little evidence linking low incomes to low growth. The poorest countries' growth rates did not significantly differ from middle-income countries between 1980 and 2001.

2. Some 11 countries out of the 28 poorest countries in 1975 were not amongst the lowest income countries in 1950 – again belying a clear link between low income levels and poor growth performance.

3. There is much stronger statistical evidence that in fact poor government causes low growth rather than initial poverty; "… stagnation of the poorest countries appears to have more to do with awful governments than with a poverty trap …"[9]

---

6   See complete list at http://www.un.org/millenniumgoals/
7   See for example the *HDR* 2003, page 41

8   Easterly, "*The White Man's Burden*", pages 40 – 52
9   Easterly, page 43

On the other hand …

One of the world's most famous economists, Jeffrey Sachs, head and co-founder of the UN's Millennium Project, takes an entirely different stance. He identifies six key factors where the cycle of poverty → low investment → low income has severe anti-developmental effects[10]:

- Low level of *human capital* due to poor health care and education will lower productivity.

- Low level of *business capital* – machinery needed in agriculture, transportation and industry – decreases potential output.

- Insufficient *infrastructure* such as roads, railways, ports, telecoms … etc, limit commerce and the ability of firms to get goods to markets.

- Poor or declining *natural capital* – arable land and healthy soils for example – provide a basis for basic human needs in developing countries.

- *Public* and *institutional capital* such as a functioning and fair legal system and commercial law enables peaceful and reliable economic transactions. It also decreases parallel markets.

- *Knowledge capital* – scientific and technological know-how not only raises productivity but 'creates value' for future generations of entrepreneurs.

Equity through taxes – redistribution of income

A good tax – Smith again

An American saying pounded into me by generations of hard-headed McGees is that only two things are for certain; death and taxes.[11] Most of my forebears would thus have looked upon the heading 'A good tax' as an oxymoron (= contradiction in terms). In any case, taxation of citizens has been a source of heavy debate, disagreement and even civil war for thousands of years. Adam Smith laid down the 'Canons (= standards, rules) of taxation' in his magnum opus *The Wealth of Nations*. According to Smith, taxes should have four main characteristics:

10 Sachs, pages 244 and 245
11 I thought this was amusing until I discovered that many countries in fact have death or funeral taxes.

1. **Certainty** – those paying should know how much they are paying

2. **Convenience** – they should be easy to collect

3. **Economy** – they should be cheap to collect relative to their yield, i.e. 'cost-efficient'

4. **Equity** – the sacrifice should be equally felt by those being taxed. (Another issue is of course that the overall effects of taxation should not be counterproductive in terms of the goals of economic policies – see automatic stabilisers in Section 3.5. under 'fiscal policy'.) Smith was referring to both efficiency and equity, where the cost of collecting the tax deals with efficiency and the ability to pay deals with equity.

Horizontal and vertical equity

Equity in tax terms means 'fairness' of the taxes levied, i.e. that the sacrifice or burden should be felt equally amongst those paying. **Horizontal equity** is 'treating equals equal' – for example when IB students get the same amount of time to complete their exams, or workers of equal experience and training are paid the same regardless of sex or age. This concept of 'equality for equals' would also apply in tax levies, where two people having the same income should pay the same tax.

Vertical equity involves 'treating different people differently' in order to enhance 'fairness'. Continuing with the examples given above; a student with a writing disability can be granted the right to 15 extra minutes in exams and minority groups can be given preferential treatment in job applications. In both these examples, the different treatment of different people can help to even out inequities. In applying vertical equity to taxes, less tax would be paid by low income earners while more tax would be paid by high income households. You no doubt realise that any form of the term 'equity' applied to tax rates is highly normative in nature.

Direct taxation

When a wage earner receives his/her wages, *income tax* has for the most part already been deducted. Other taxes which are levied directly on individual incomes are taxes on profits, interest received, capital gains taxes (on income earned by selling property or shares) and dividends from the ownership of shares. Other economic agents, e.g. firms, pay *corporate taxes* (often called profit taxes), and *labour taxes*. In each of

MACROECONOMICS

377

these cases, the tax is clearly distinguishable and goes directly from the taxpayer to the tax office; these are **direct taxes**. (Refer back to Section 2.2; PED and taxation, 'Sorting out the terms; different forms of taxation'.)

Direct taxes on income are associated with two main economic effects.

1. The first is the *redistribution effect*, whereby income tax is collected and then redistributed to other (less fortunate) members of society. Note that this is not only in the form of money (see 'transfer payments' below) but also in the form of health care, education and road networks ('benefits in kind' below).

2. The second effect is the possibility of a *disincentives effect*; when taxes on income increase at higher income levels, workers might not view additional working hours as worthwhile. It is also possible that an unemployed person gets a job and incurs a net loss of disposable income when income tax is paid at the same time as various social benefits disappear – this is a form of *poverty trap* for low income households. In addition, there is the possibility of a **black labour market** when increasing tax levels create an incentive for workers to avoid taxes by not reporting income to the tax office.

Indirect taxation

One of the snappier comebacks by one of my students in high-tax Sweden was a retort to my inevitable chastisement 'My tax money paid for that!' when pupils complained about the school lunch. 'My tax money paid for it too! Oh, and also for your salary.' Cheekiness aside, the point is well taken; a good deal of government tax revenue is comprised of taxes on expenditure such as value added taxes (*VAT*), specific taxes commonly levied on petrol and alcohol (*excise duties*) and import taxes (*tariffs*). All of these are **indirect taxes**, since economic exchanges such as consumption/expenditure rather than individuals are taxed. The tax paid is baked into the exchange and goes – indirectly – to government via the firms selling the goods.

Indirect taxes affect supply which implies that market equilibrium is negatively affected; the supply curve for the good shifts left. While the case is often that this causes a misallocation of resources (and deadweight loss), we have also seen that in fact taxes might serve to decrease negative externalities and therefore instead increase allocative efficiency. An issue worthy

of notice here is whether a tax on a good having negative externalities should be designated to cover only its own costs or not. For example, many countries have road taxes which contribute to government tax receipts far in excess of what is subsequently paid for building and renovating roads. A strong case can be made by car owners that they are paying more than their fare share of taxes, since any surplus receipts will benefit those who do not own cars. The counterargument is that road usage is strongly associated with negative externalities such as pollution and noise, so the additional tax is an adequate disincentive for road use.

> **Definition: 'Direct and indirect taxes'**
>
> **Direct taxes** are levied on economic agents' income, wealth or property. Firms pay profit tax and labour tax. Households pay income tax, capital gains tax and property tax.
>
> **Indirect taxes** are levied on consumption and expenditure. Value-added tax, excise duties (special taxes on tobacco and alcohol) and tariffs (taxes on imports) are examples of indirect taxes.
>
> **(WARNING!** Many students confuse tariffs with excise duties. Perhaps it is because 'duties' is so easily associated to 'Duty Free'! Whatever the origin of the confusion, an excise duty is a tax on 'bads', e.g. alcohol, tobacco and petrol. Oh, one of my American students informed me that there is often an excise duty on gambling in the US. I looked it up – he's quite right.)

Progressive tax

Direct taxation has one clear advantage over indirect taxes, namely that a direct tax can be adjusted to conform to societal views on equity. Income tax rates can be adjusted to each person's ability to pay, i.e. adjusted to income. A **progressive tax** on income means that higher income will result in a higher percentage of tax paid, i.e. an increasing proportion of income goes to tax. Most countries will have a systematic increase in the proportion (= percentage) of income tax paid as income rises, since this is virtually the only way in which income can be redistributed – by 'taking from the rich and giving to the poor'. Commonly in income tax systems, there is a minimum income level where no tax is paid, whereupon the marginal tax – the tax paid on the last money earned – increases.

For example, say that income tax on the first €2,000 is zero but 15% on any income above this. Earning €3,000 would mean that income tax would be paid only on the additional €1,000 – the amount exceeding the threshold of €2,000. This is the *marginal tax rate*. Tax at an income of €3,000 would be €1,000 × 0.15 = €150. (However – take heed! – the *average tax* on income is of course total tax paid over total income; €150 / €3,000 = 5%.)

The progressive taxation element in this method of income taxation is that higher income brackets will mean higher percentage tax paid. Continuing with the example, say that the tax rate progressively increases to 20% for income above €5,000 but below €10,000, and that a person's income increases from €5,000 to €7,000. The **marginal tax** on the €2,000 above the €5,000 tax bracket is €400 while the *average tax* paid will be €3,000 × 0.15 + €2,000 × 0.2 = €850. The tax rate then increases at every higher income bracket. This is illustrated in the upward sloping – progressive – curve in Figure 55.7, where the marginal tax rate is of course the slope of the curve.

Proportional taxes

A **proportional tax** is exactly what it sounds like; a percentage of income paid in tax. Since the percentage is unchanged at higher income levels, there is no marginal tax effect and any rise in income will add to total tax payment at a constant rate, so average tax rate is unchanged. In other words, the proportional tax curve will have a constant slope, as illustrated in Figure 55.7. Capital gains, corporate profits and dividends are types of income which are frequently taxed on a proportional basis.

Regressive taxes

Just as indirect taxes can be 'flat rate' – such as unit taxes on wine – direct taxation can consist of a fixed sum which does not change as income rises, which means that average taxes paid fall as income rises. A **regressive tax** means that the average proportion of tax paid on income or profit falls as income/profit increases. For example, a yearly business registration tax of £1,000 for a small corner shop with £20,000 in profit means 5% average tax. For Imperial Tobacco Group PLC, the £423 million in profit in 2002[12] would mean that the registration tax is on average 0,00023%. Since the average tax payment as a proportion of income is falling, the regressive tax curve will become successively shallower, shown below in Figure 55.7.
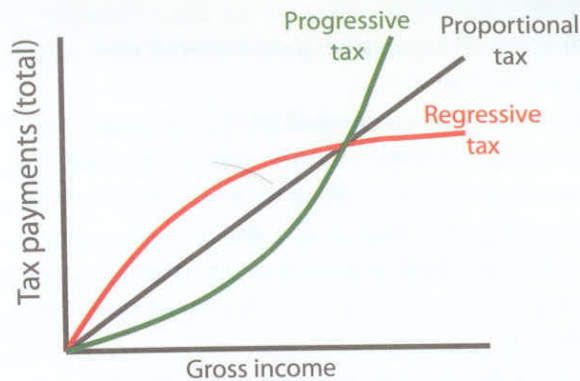

Figure 55.7 Progressive, proportional and regressive taxes

Another regressive effect of taxation is that lower income groups are often hit harder by *indirect taxes* than higher income groups. Alcohol, tobacco and petrol are all major contributors to total government tax revenues and contain a large element of *flat-rate tax* (excise duty). The regressive element herein is when low income earners pay the same tax as high income earners; the former are spending a greater proportion of income on the goods than the latter. In fact, value added taxes are arguably regressive since a 10% sales tax on a £100 purchase constitutes a far larger part of income for a person earning £12,000 than for someone earning £120,000. One could say that this lowers the effectiveness of any intended redistribution effects of the tax.[13]

---

12 Annual report Imperial Tobacco PLC at http://www.imperial-tobacco.com

13 I frequently put forward scathing comments on how corrupt police in Mexico are in fact a symptom of regressive taxation. When police stop drivers for real or made-up offenses, they have a 'price list' of fines at the ready (they can draw that list faster than Wyatt Earp could pull his six-shooter). They then make clear how much time, effort and inconvenience it will be for the driver to have to go through the motions of paying the official fine – often claiming that they will also have to confiscate the vehicle. They are waiting for the driver to say "So, how can you help me officer?!" A common bribe ('mordida' – roughly 'little bite') is 200 Mexican pesos, about USD18. This is four days' earnings for about 40 million people here in Mexico and about 25 minutes for an expatriate IB teacher. In other words, a flat-rate bribe will hit Benito the day labourer far harder than Graham the IB coordinator. Oh yes, before I forget; the State of Mexico toughened the driving laws in 2007, increasing the number of possible violations while driving. My Mexican friends rather fatalistically called the harsher legislation a tax raise. I called it the cheapest possible form of pay increase for the police – it gives the police more power to extort bribes and the government pays nothing.

MACROECONOMICS

> **Definition: 'Progressive, proportional and regressive taxes'**
>
> **Progressive** tax means that the percentage of tax paid increases as income rises – a larger proportion of income goes to tax as income rises.
>
> When the percentage paid is the same no matter what income level, the tax is **proportional** – the same proportion of income goes to tax regardless of income.
>
> A tax that is the same amount of money (flat-rate) regardless of income is a **regressive** tax – a decreasing proportion of income goes to tax as income rises.

Transfer payments and services in kind

Societies often find it necessary and prudent to provide benefits for certain groups of citizens. Social (welfare) benefit systems which redistribute income via cash **transfer payments** exist in most countries in some form. Students will receive grants and soft loans; retired people receive pensions and additional health care monies; low income households and single parents (far too often one and the same) receive supplemental housing allowance and welfare payments; and unemployed people receive unemployment benefits and perhaps travel contributions for job seeking.

The other main form of societal redistribution is **services in kind**; merit goods such as health care and education. Since these goods would be both underprovided and thus underconsumed on a free market, it would be the poorer groups who would suffer the most. By using government (tax) monies to provide these goods on a general basis of not-for-profit, total economic welfare is increased.

> **Definition: 'Transfer payments'**
>
> Any government monies granted to households – without some form of corresponding output – is a transfer payment. Unemployment benefits, housing allowances for low income families and social benefits are examples of transfer payments.

> **Definition: 'Services in kind'**
>
> Under-provided and under-consumed goods such as public transport, education and health care are services often provided directly or indirectly (subsidies) by governments. These **services in kind** have highly positive regressive effects as they benefit poorer households proportionally more than wealthier.

Evaluation of equity policies

All taxes will in some way lead to market distortions, since they lead to a decrease in supply and concomitant decrease in quantity demanded. Taxes on labour lead to higher costs for firms and an increase in unemployment; indirect taxes lead to less goods being consumed; corporate taxes lead to fewer new firms and less investment; and capital gains tax reduces incentives and lowers economic activity. However, the economic argument in favour of enduring these distortions is that the allocative losses can be made up for by the overall redistribution gains to society in terms of public and merit goods. Governments use tax receipts to provide goods which are beneficial to all of society. For example, the efficiency loss in firms due to labour taxes are offset by productivity gains arising from a well educated work force, and the relatively small decrease in consumption arising due to high marginal taxes on the richest 1% can be offset by welfare gains when the taxes are redistributed to poorer groups.

And justice for all ...

It gets trickier when governments have to decide which tax creates the 'fairest' outcome. Indirect taxes have the advantage of being simple and easy to collect, but have the disadvantage of levying a heavier burden on the poor than the rich. For example, a unit tax of €2 per litre of alcoholic beverage will have a far greater real income effect on Otto Normalverbraucher[14] purchasing a €3 six-pack of beer than it will have on Countess Antoinette du la Monet[15] buying a €150 bottle of 1966 Chateau Neuf du Pape. In effect, a flat-rate tax expenditure tax will have strong regressive tax effects on poorer groups.

As for income taxes, where is it written that higher income must lead to higher proportion of tax – when in fact a proportional tax already means that higher income leads to more tax payments?! The argument for progressive income taxes ('higher income = higher percentage tax') is highly normative, in that there is an evening-out effect of incomes which is 'fair' to society in

---

14 German for 'Joe Average'
15 Pronounced 'de Money'

general. Increased income equality also has economic benefits, such as lower crime rates and inner city regeneration which, of course, benefits everyone. Quite naturally, the well-off point out that between one third and half of all income tax receipts in developed countries come from the top 5% income earners – the implication being that this is 'more than their fair share'. Another common argument put forward by the affluent is that progressive taxes are a major disincentive for people to push themselves, being in fact a 'punitive tax' on achievement, hard work and entrepreneurial spirit.

Normative economics – link to development issues

Finally, the field of development economics is rife (= common) with a number of additional key concepts which are either very broadly used and/or subject to the same **normative limitations** as above. Using terms such as 'equitable' or 'inequitable' distribution of income/resources clearly means that one has drawn a normative line somewhere. Just imagine two societies; Society A consists of 1 million people where 980,000 people are well off and 20,000 are living in poverty and Society B of 2 million people where 30,000 live in poverty. Which society is 'best' in terms of income distribution? Well, if you use 'amount of poor people' then A is preferable. If you use 'percentage living in poverty' then B is likely to be preferable. Now, assume a third society, C, consisting of 30,000 people where 29,999 lived in poverty. Using 'amount of poor people' to define 'best' would point to Society C being preferable to both A and B. Of course my example is hideously exaggerated, but the point is that positive economics runs into some severe obstacles in development.

Yet another trade-off?

Notice the question mark in the above heading and prepare yourself for another of my 'No, there is no answer' paragraphs. Redistributing income has a double-edged price tag of administrative costs to government and allocative losses to society. Redistribution has been likened to attempting to transfer water from one bucket to another.[16] A portion of the water – income – will invariably be lost in the process, i.e. there will be costs associated with administration and economic efficiency. The taxes on economic endeavours such as labour and investment also render opportunity costs in the form of forgone … labour and investment. This argument thus puts forward that attempting to improve social welfare by increasing

equity will render a social cost in terms of greater inefficiency in the use of resources which leads to a loss of income.

The basic question here is whether the costs to society of trying to increase equity are counterproductive in terms of economic growth. The trade-off, according to this line of reasoning, is between growth in total income and growth in income equality. In the final analysis, a good many studies show that there is indeed a trade-off – increased income inequality is 'pro-growth'. However, a good many studies show the reverse, i.e. that greater inequality leads to lower growth! And for my final trick; a recent study by the OECD showed "… no evidence that the level of income inequality affects GDP one way or another."[17] I told you there would be no answer.

---

17 *OECD Economic Studies No. 35, 2002/2* at http://www.oecd. org/dataoecd/42/33/22023319.pdf

---

16 Expressed by the famous economist *Arthur Okun* in his book *Equality and Efficiency: The Big Trade-off*, Washington DC, Brookings Institution, 1975.

HL extensions

Calculating the average tax rate

The table in Figure 55.8 outlines three income brackets of between £10,000 and £200,000. Assume that taxes are levied at a regressive, proportional or progressive rate. The flat rate and proportional taxes should be no problem filling in.

Figure 55.8 Summary of income taxes

	Tax payments on a gross income of:					
	£10,000 – 20,000		£20,001 – 40,000		£40,001 -	
	Tax paid (£)	Average (%)	Tax paid (£)	Average (%)	Tax paid (£)	Average (%)
Regressive tax (flat rate £1,500)	£1,500	xx%	£1,500	xx%	£1,500	xx%
Proportional tax (15%)	£xxx	15%	£xxx	15%	£xxx	15%
Progressive tax (rates of 15%, 30%, 40%)	£xxx	15%	£xxx	20%	£xxx	30%

Using the *progressive* income tax tiers, calculate the total income tax paid by Bob (gross income of £19,000), Lisa (£35,000) and Leslie (£75,000).

Calculating the marginal tax rate

Progressive taxes mean a higher percentage tax paid on higher incomes. (In Sweden we say 'the tax on the last SEK100 earned' if this helps.) Thus, for incomes up to £20,000 the tax is 15% but any income in the next two tiers will have a higher average tax rate. The marginal rate is calculated by taking the change in total tax paid divided by the change in income, times 100; ($\Delta T/\Delta y$) x 100. Note that the marginal rate will stay the same moving from £12,000 to £14,000 within the same tax tier (15% on the additional £2,000) but will be higher when moving into the next tier.

- Bob's gross income increases from £19,000 to £21,000. Calculate the marginal tax rate on the additional £2,000. (See footnote for some help.)[18]

---

18 Bob is initially paying 15% on £19,000; £2,850. At £21,000 he will pay 15% on £20,000 and 30% on the £1,000 in the higher tax bracket. Add these two values and you get the *new average tax* paid. Deduct original average tax paid (£2,850) and you have the numerator in the marginal tax formula. Stick in the denominator (change in income of £2,000) and multiply by 100.

- Calculate Lisa's marginal tax when her gross income increases from £35,000 to £44,000.

- 12,000 is 1,800 …. and 14,000 is 2,100 …. increase in tax is 300 and increase in y is 2,000; this is a 15% tax. $\Delta y = 2,000$ … new tax is (20,000 × 0.15 + 1,000 × 0.3 = £3,300)

- Old tax is 2,850

- ΔT; 3,300 – 2,850 = 450

- $\left(\dfrac{450}{2,000}\right) \times 100 = 22.5\%$

Summary & revision

1. **Equity** is the normative concept of 'fairness' in the distribution of wealth and income.

   a. **Horizontal equity** means 'treating everyone the same' – everyone over 18 gets a vote and nobody can be denied entrance to university based on sex or income.

   b. **Vertical equity** is treating some individuals differently to create 'fairness' – such as a parliamentary quota system for women (50% of MPs in Norway must be women) and preferential university entrance for minority groups.

2. **Equality** means equal shares of wealth and income.

3. Income inequity is often seen as a **failure of the market system**.

4. The **Lorenz curve** measures inequality of income distribution by plotting out the cumulative percentages of income and cumulative percentages of households. The line off perfect equality means that 1% of households receive 1% of income, 5% of households receive 5% of income … etc. The further away from the line of perfect equality a country's Lorenz curve is, the more unequal the distribution of income.

5. The **Gini coefficient** is based on the Lorenz curve. It measures the area between the Lorenz curve and the line of perfect equality as a proportion of the total area under the line of perfect equality. Values can range from 0 to 1 where the higher the value the worse the distribution of income.

6. **Poverty** has many causes – primarily the effects of low incomes. Low savings rates and investment lead to low income (poverty trap) which perpetuates low levels of governments spending on merit goods and infrastructure … and thus low incomes.

7. **Effects of poverty** are low standards of living and lack of opportunities. This means malnourishment, poor education, low paying jobs, poor health, low life expectancy and destruction of natural resources.

8. **Direct taxes** are taxes going directly to government from the taxpayer, for example income tax, profit (corporate) tax and capital gains tax (tax on profits made selling shares or a house).

9. **Indirect taxes** are taxes based on a *transaction* and go from taxpayers (e.g. consumers) via firms to government – expenditure taxes such as ad valorem (value-added) taxes and excise duties are indirect taxers.

10. A **progressive** tax means that the average *proportion of tax paid increases* as income rises. **Proportional** tax has the *same percentage* average tax at every level of income. **Regressive** tax means that the *average tax decreases* as a percentage of income decreases as income rises.

11. Governments commonly intervene in markets to **increase equity** via various forms of *income re-distribution*. Common methods are *taxes on luxury goods*; *subsidies* for basic necessities such as milk and rice; hugely *differentiated property taxes*; *marginal tax rates*; various forms of *transfer payments* (social benefits and housing allowances for example); and *services in kind* (public and merit goods whose benefits are highly regressive – lower income groups benefit proportionally more than higher income groups).

HL extension

12. The **average tax rate** is calculated as the average tax paid divided by total income times 100

$$\left(\left[\frac{T}{y}\right] \times 100\right)$$

13. The **marginal tax rate** is calculated by dividing the change in tax paid by the change income times 100

$$\left(\left[\frac{\Delta T}{\Delta y}\right] \times 100\right)$$

MACROECONOMICS

2.4

56. The Government Budget

Key concepts:

- Government revenue and expenditure
- Budget outcome – surpluses, deficits and debt

Alexander Hamilton started the U.S. Treasury with nothing – and that was the closest our country has ever been to being even.

Will Rogers

Government revenue and expenditure

Just like you might make a list for expected inflows of money and planned spending for the coming semester, governments must do the same. (And, just like you, governments often spend more than they receive.) A **government budget** is basically an outline or proposal of all government spending, transfer payments and tax receipts for the coming fiscal year.[1] The budget lists all incoming **tax revenue** (often called tax receipts) and other inflows (such as profits from government firms and various fees and licenses paid by households) and also all outlays for **government spending**. This can be divided into *current spending* (goods and services such as social services, health care and education); *capital spending* (government investment into roads, bridges and telecommunications); and *transfer payments* (social security and government run pension schemes).

It is worth noting that both the composition and amount of spending is quite politically motivated, in fact, the budget is often something of a political agenda in itself since it mirrors ideologically motivated spending, for instance on defence and social safety nets.

Revenue	Expenditures
+ taxes (income taxes, VAT, profit taxes…etc) + profits from state-owned enterprises + social security contributions + sales of state-owned businesses, e.g. privatisation	- defence - education - public works - health care - central, regional, local government - social security - roads - bridges - …etc…
+100	−100

Figure 56.1 Illustration of government surplus and deficit (over a cycle)

Budget outcome – surpluses, deficits and debt

There are 10^{11} stars in the galaxy. That used to be a huge number. But it's only a hundred billion. It's less than the national deficit! We used to call them astronomical numbers. Now we should call them economical numbers. Richard Feynman.

---

1   A 'fiscal' year, as opposed to a calendar year, is a **chosen time period** for 'making ends meet' and can run from August to August, October to October or any other 12 month period. The fiscal year often coincides with the deadline for parliament/congress accepting the budget.

Figure 56.1 is a highly stylised illustration of a government budget – and the large illustration is also heavily idealised since it portrays a perfectly balanced budget. (Note that I have not bothered to include currency units.) When government receipts exceed expenditure there is a budget surplus; when expenditure exceeds receipts a budget deficit. This is shown in the two smaller figures, which in a Mermaidomics (= highly imaginary) economic cycle *balance out*; the surplus during a boom period is used to fund deficit spending during a recessionary period. In reality government spending often exceeds receipts over a substantial portion of the business cycle and the government is forced to *borrow* in order to finance the deficit.

Let's put this into a simple example. Assume and economy with an initial balance budget (Figure 56.2) and no borrowing or initial debt. Total government revenue of $150 equals total expenditure during fiscal year 1. During the next fiscal year, the government borrows $20 to be paid off over a ten year period at a fixed interest rate of 8%. This benevolent government uses the loans to increase services throughout the next four year period (which, oddly enough, is the same length as a US presidential period) so total revenue equals the outlay for services. However, paying of 10% of the principal (= original amount borrowed) every year together with interest of 8% means that the government now has to service its debt – seen in the debt servicing column.

The country increases borrowing at an even rate and spends all of this on additional societal services. This leads to increasing debt servicing and increasing deficits – not to mention that total debt increases rather drastically. Here is how the debt servicing and total debt was calculated:

Fiscal year 2: loan of $20 means a payment on principal of $2 plus interest of 8% on $20 (0.08 x $20); **$3.6**

Fiscal year 3: another 10 year loan of $30 plus remaining loan of $18 (recall that 10% of the principal of the year 2 loan was paid off) means an outlay to pay off the principal loans of $3 plus $2. Together with 8% interest on loans ($18 x 0.08 + $30 x 0.08) we get $5 plus $3.84; **$8.84**. This borrowing yields a budget deficit of the same amount.

Figure 56.2 Example of the link between deficits and debt

	Tax rev	Other rev	Loans	Tot rev	Services	Debt servicing	Tot exp	Balance	Tot debt
Fiscal year 1	100	50	0	150	150	0	150	0	0
Fiscal year 2	110	60	20	190	190	3.6	193.6	-3.6	18
Fiscal year 3	120	70	30	220	220	8.84	228.84	-8.84	43
Fiscal year 4	130	80	40	250	250	15.64	265.64	-15.64	74
Fiscal year 5	140	90	50	280	280	23.92	303.92	-23.92	110

Yes, there's a moral here; borrowing today means debt tomorrow. Even though our very benevolent government in the example here is using the increase in borrowed funds for schools and day care centres (rather than, say, nuclear weapons) the debt servicing leads to a final deficit of $23.92 and debt of $110 during fiscal year 5. The sum of this borrowing over time, known as the **national debt**, can severely hamper (= hinder, obstruct) governments' future fiscal freedom in two ways:

1. The government gets a reputation for being a '**high risk debtor**' and will have increasing difficulties in getting new loans. How does a high-risk borrower get a loan? Simple; he/she will have to pay a higher price, i.e. interest rate. This limits governments spending options.

2. Increased debt servicing naturally means higher outlays for government. Debt servicing costs yield rather hefty opportunity costs in terms of foregone roads and school books.

Income

- Income Tax — £150
- National Insurance — £99
- VAT — £81
- Excise Duties — £46
- Corporation Tax — £43
- Council Tax — £25
- Business Rates — £25
- Other — £79
- Borrowing — £149

Expenditure

- Social Security — £195
- Health — £122
- Education — £89
- Debt Interest — £44
- Defence — £40
- Public Order — £35
- Personal Social Services — £32
- Housing & Enviornment — £27
- Transport — £22
- Industry, agriculture & employment — £20
- Other — £73

(Source: HM Treasury at http://www.hm-treasury.gov.uk/budget2012_documents.htm)

Figure 56.3 UK government deficit, fiscal year 2010 – 2011

Definition: 'Government budget, deficit, surplus, national debt, foreign debt'

The government **budget** is a fiscal policy instrument and political intention comprising government spending (based on receipts) for a fiscal year.

When government receipts are higher than spending there is a **budget surplus**.

When government spending exceeds receipts there is a **budget deficit**.

The sum of deficits minus surpluses over time equals the **national debt**.

The portion of the national debt which is owed to foreign banks/governments is the **foreign debt** (or **external debt**).

Figure 56.3 shows how the UK government went into severe deficit spending during the fiscal year 2010 – 2011, chalking up a massive £149 billion in debt. This borrowing accounted for *over 21% of all government spending* and shows the severity of the impact of the economic crisis on the UK economy.

Summary & revision

1. The government budget is a proposal of all expenditure and revenues over a period known as a fiscal year.

2. Government **revenues** consist primarily of tax revenues and profits from government.

3. **Expenditure** is comprised of:

    a. *Current spending* - social services, education, defence

    b. *Capital spending* – infrastructure, schools, hospitals

    c. *Transfer payments* – pensions, social security payments

4. When revenues exceed expenditure, the government is running a **budget surplus**.

5. A government **budget deficit** is when expenditure is greater than revenues. The government borrows money to fill the gap – this is the government **budget deficit**.

6. The cumulative value of past deficits is the **national debt**.

57. The Role of Fiscal Policy

Key concepts:

- Fiscal policy – government expenditure and taxes – effect on AD
    - o Mind the gap – inflationary and deflationary gaps and fiscal policies
    - o Effects of expansionary and contractionary fiscal policy
        - Keynesian model
- Automatic stabilisers
- Fiscal policy and LR growth (AS … interventionist S-side policies)
- Evaluation of fiscal policy
    - o Advantages of demand-side policies
        - Target sectors of the economy
        - Effects on AD
        - Promoting economic activity during recession
    - o Disadvantages of demand-side policies
        - Trade-off problems
        - Time lags
        - New-classical critique
        - Four key points
        - Political constraints
        - Limits of fiscal policy – S-side shocks cannot really be dealt with

"An economist's guess is liable to be as good as anybody else's." Will Rogers

The four 'mainstream' macro objectives and possible monetary and fiscal policy responses to various stages in the business cycle are given in Figure 57.1. Notice, once again, that any use of monetary and fiscal policies aimed at demand management will have noticeable trade-offs, where adjusting aggregate demand in order to achieve a particular macro objective will have negative effects on some other objective. As briefly outlined in Chapter 48, it is quite evident that many of the macro objectives are very difficult to attain simultaneously, leading to conflicts and opportunity costs for the economy.

Figure 57.1 Macro objectives and possible policy responses

	Overheating economy (inflationary gap)	Recessionary economy (deflationary gap)
Macro objectives		
1. High/stable growth	High growth rate	Falling growth rate or negative growth
2. Stable prices, i.e. low inflation rate	High or increasing inflation rate	Low(-er) inflation rate, possibly deflation
3. Low level of unemployment	Low unemployment	Increasing/high unemployment
4. Trade balance (X = M)	M often larger than X; trade deficit	M falling; possible trade surplus
Possible **fiscal** policies	ΔT, ΔG/transfer payments	ΔT &/or ΔG/ transfer payments
Possible **monetary** policies (Chapters 58 and 59)	Tight monetary policy; ΔSm, Δr	Loose monetary policy; ΔSm, $\Delta^- r$

	Overheating economy (inflationary gap)	Recessionary economy (deflationary gap)
Positive effects of policies	Lower inflation, relieves pressure on tight labour market, possible improvement in trade balance	Decrease in unemployment, increased output (or slower negative growth), societal benefits
Negative effects of policies	Growth rate falls back, lower investment can harm long run potential growth	Inflationary pressure, increase in imports might cause trade deficit

Fiscal policy – government expenditure and taxes – effect on AD

Governments can influence aggregate demand in the economy by using taxes (T) and/or government spending (G).

An increase in **income taxes** lowers households' disposable income which in turn lowers consumption and aggregate demand. Increased **expenditure taxes**, e.g. VAT, have the same effect due to the decrease in real income.

- $\Delta{\uparrow}T \to \Delta{\downarrow}C \to \Delta{\downarrow}AD$ … or … $\Delta{\downarrow}T \to \Delta{\uparrow}C \to \Delta{\uparrow}AD$

Since the government has the power to adjust government spending – which is a component of aggregate demand – this will have a direct impact on total expenditure in the economy. A boost in **government spending** will lead to an increase in aggregate demand and vice versa.

- $\Delta{\uparrow}G \to \Delta{\uparrow}AD$ … or … $\Delta{\downarrow}G \to \Delta{\downarrow}AD$

The automatic stabilising effect of taxes and social benefits (see below) are enhanced by the intentional adjustments of government spending and/or changing taxes, known collectively as *discretionary fiscal policies* as they are 'at the discretion' of government.

- During a *recession*, the government increases spending on roads, education and such, and this has the effect of increasing aggregate demand – of which government expenditure is a component – and taxes can also be adjusted downwards in order to increase disposable income and induce increase consumption. This is called **expansionary fiscal policy**.

- When the economy shows signs of *overheating*, government can cut back on spending and increase taxes in order to lower aggregate demand. The aim is also to even-out the business cycle, as in Figure 57.2, bringing cyclical variations closer to the long term trend. (The extent to which this actually works is hotly debated and we will look at this briefly under the heading 'Time lags' below.) These are **contractionary fiscal policies**.

Fiscal policy was the main instrument used during the 1950s and '60s to influence aggregate demand, where Keynesian theory prescribed using government spending as a 'rudder' to adjust the economy. According to the Keynesian view of demand management during the 20 year period following WWII, the focus should be on lowering unemployment. Deficit spending could be used during recessions as a way of increasing output/growth and thereby increased employment. The aim was to use 'good times' to even out 'bad times'; government surpluses created by low(-er) government spending and higher tax receipts during booms were to be used to stimulate the economy during recessions, thus evening-out the business cycle and creating long run stability. During the 1970s, a number of things – most noticeably increasing government debt and both high inflation and high unemployment – caused many governments to abandon attempts to fine tune the economy and look at alternative policies, most noticeably supply-side measures.

Mind the gap – inflationary/deflationary gaps and discretionary fiscal policies

Figure 57.2 illustrates how the government can influence economic activity by using fiscal and monetary policies. This 'cutting the peaks and filling the troughs' (Diagram I) by way of adjusting aggregate demand is commonly undertaken in order to cool down an *overheating economy* (at around t_0 and beyond) and to pull the economy up out of *recession* (at around t_1 and beyond). Using fiscal and monetary policies to adjust aggregate demand and even-out business cycle fluctuations is often referred to as *fine tuning*. It was the main form of economic control exercised by governments during the 1950s and 1960s in industrialised market economies. It has largely fallen from use as a 'fine tuner' due to the problems associated with *time lags* – see further on.
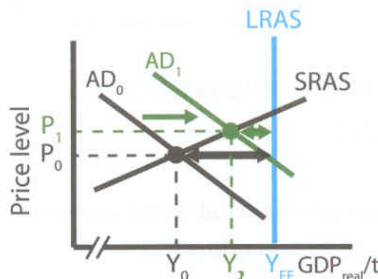
I: The business cycle and demand management
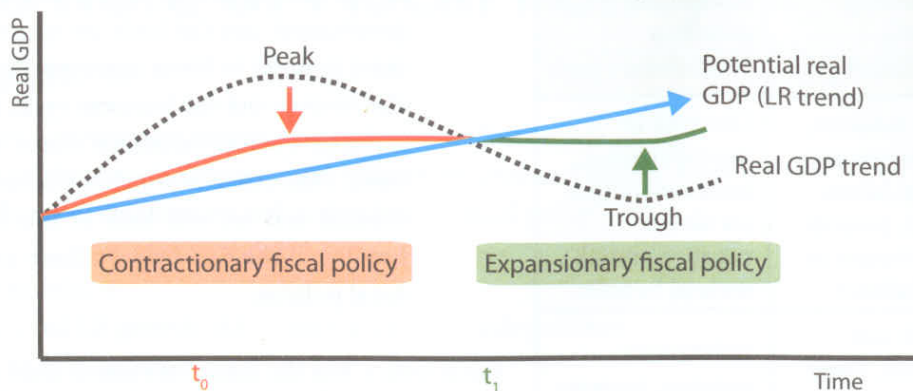
II: Contractionary fiscal policies and AD

III: Expansionary policies and AD
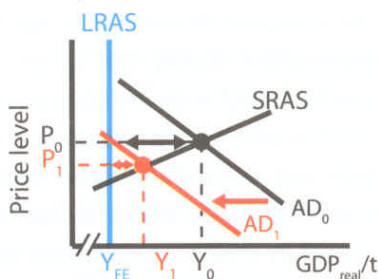
Fiscal policies: increased taxation, lower government spending (notice that the new level of AD closes some of the inflationary gap)

Fiscal policies: decreased taxation, increased government spending (the increase in AD closes the deflationary gap to a certain extent)

Figure 57.2 I, II and III; Demand management, business cycle and AD

- **Cooling the economy:** When real GDP increases beyond the long run trend of potential GDP (t_0 in the business cycle diagram) there are negative influences on several of the main macro objectives, such as rising *inflation* and possibly a *trade deficit* (exports < imports). The government can try to countermand this by **contractionary fiscal policies:**

  o Government can *lower government spending* and/or ...

  o ... *increase various taxes* in order to reduce consumption and/or investment. As C and I are components of aggregate demand, the AD curve decreases from AD_0 to AD_1 in **Figure 57.2 II.**

- **Stimulating the economy:** An economy in recession will warrant stimulatory demand management policies, **expansionary fiscal policies:**

  o When the economy is operating below potential GDP and moving into recession, around t_1 in **Figure 57.2:I**, government can lower taxes and/or increase government spending. Lowering income taxes would increase consumption and government spending is an AD component. Aggregate demand increases from AD_0 to AD_1 in **Diagram III.**
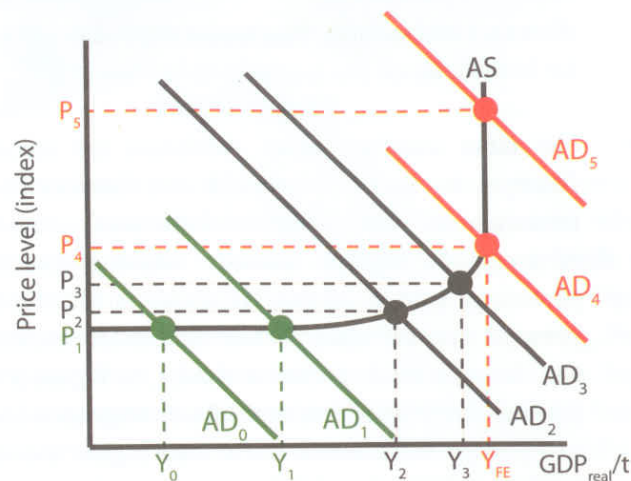
If successful, the fiscal policies will serve to keep the economy 'in line' with long run potential real GDP, shown by the business cycle's lower amplitude around the trend line and the economy moves closer to the long run equilibrium shown in **Figure 57.2:I.**

Effects of fiscal policies – Keynesian model

I have used the new-classical model above to illustrate the effects of fiscal policies on inflationary and deflationary gaps. The Keynesian model yields other possibilities and insights, as seen in Figure 57.4.

Four possible outcomes of fiscal policies arise depending on where macro equilibrium is initially:

1. Low levels of income and high unemployment (Y_0) would give rise to expansionary fiscal policies (AD_0 to AD_1) which would increase income without inflation.

2. Approaching full employment (Y_{FE}), firms encounter bottle-necks in supply and diminishing returns – costs rise. Fiscal stimulation at Y_2 or Y_3 will increase income and decrease unemployment but the trade-off is increased inflation (P_2 to P_3) in the economy.

3. Fiscal stimulus at Y_{FE} will be purely inflationary as the economy is utilizing all factors of production maximally and any increase in AD has no effect on real GDP.

4. Deflationary policies undertaken at P_4 and Y_{FE} will decrease inflation without causing a decrease in real GDP or a rise in unemployment (AD_5 to AD_4).

The shape of the Keynesian aggregate supply curve has some rather important ramifications for fiscal policy. The first is that government can 'choose from a menu' of possible inflationary and real income levels. The second is that governments must accept that there will be trade-offs in setting an unemployment or inflation target. And finally, as will be seen below, Keynesian theory puts forward that supply-side policies are ineffective at very low levels of income (say at Y_1 or below) since the increase in AS does not have any effect on equilibrium income.

At the full employment level of output, any **expansionary fiscal policy is purely inflationary.**
Conversely, any **deflationary policy has no effect on income**.

Approaching the full employment level of output (Y_{FE}), any fiscal policy – either expansionary or contractionary – will have a **trade-off effect** in terms of income and inflation (or unemployment and inflation).

At low levels of income it is possible to implement expansionary fiscal policies **without causing inflationary pressure**.

Figure 57.3 Fiscal policies and the Keynesian AS curve

MACROECONOMICS

Automatic stabilisers – safety nets

One of my many childish allegories is that the economy is like a sailing ship and I inflict this on my students as the need arises. Like right now. Picture a ship sailing along on the seas; waves and wind exert forces on the ship's direction and stability and may force the ship towards reefs and rocks. The stability and direction of the ship is set by two primary forces; the keel and ballast which serve to *automatically* stabilise the ship. The rudder which is in the hands of the captain and at his *discretion* (= will, choice, freedom) can also be used to help stabilise the ship and plot a chosen course.

When the economy starts to 'heat up', a number of stabilising effects will automatically kick in – mechanisms which are built-into the economic system and social welfare system.

- Government spending in the form of **social benefits** is one such stabiliser. An increase in economic activity and GDP will see lower unemployment levels and serve to lower the need for social benefit payments such as unemployment benefits, income-based housing

allowance and so forth. This lowers disposable income for households on the receiving end of benefits.

- The other stabiliser, **taxes**, influences net income (income after tax) of households and therefore their consumption. Most countries have some element of *progressive* income taxation, meaning that the percentage paid in income tax increases when gross personal income rises. When household incomes rise during a boom period, around t_0 in **Figure 57.4 I**, either due to overtime, new jobs or wage increases, many wage earners will move into higher income brackets and thus pay a larger proportion of their take-home pay in tax.

Taken together in the macro environment, lower social benefits and higher proportionality of taxes paid both serve to lower net disposable income in the economy and therefore consumption. Since both social benefits and progressive income taxes are built into the economic and social system, the braking effect on aggregate demand, shown by the downward readjustment from AD_1 to AD_2 in **Diagram II**, is automatic. The effect is that aggregate demand falls less than otherwise would be the case, cutting the peak of the business cycle.
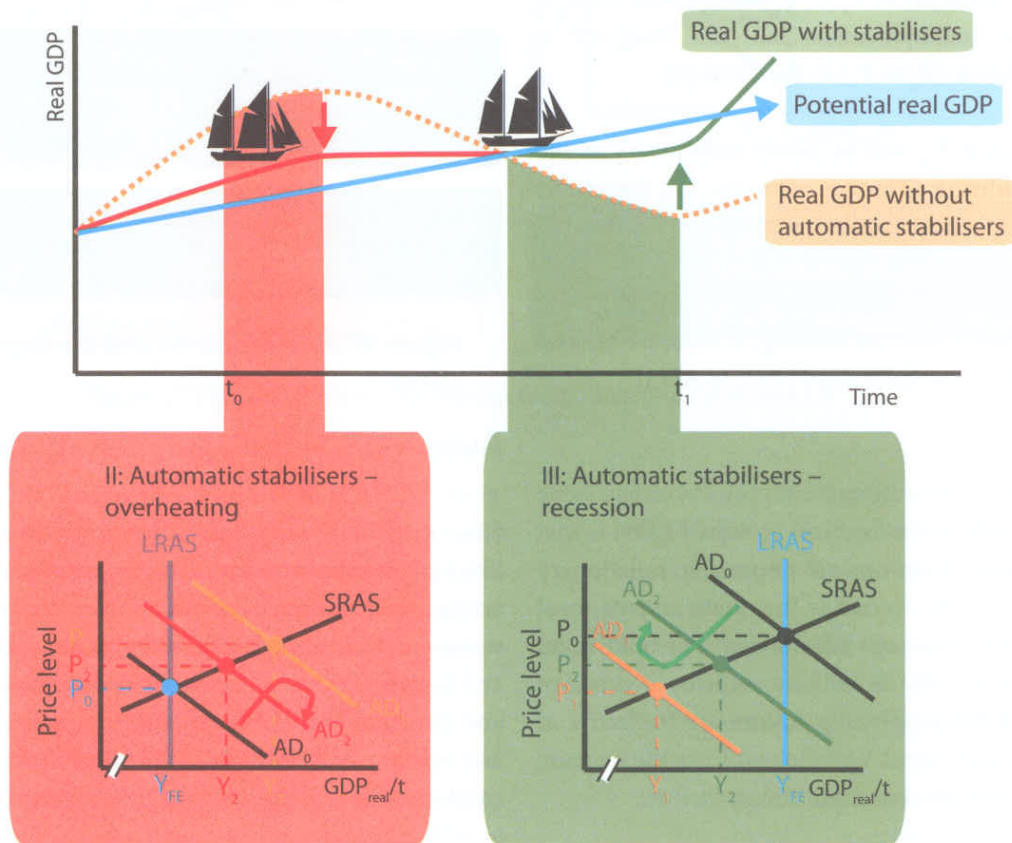

Figure 57.4 I, II and III: Business cycle and the effect of automatic stabilisers on AD

During **recession**, around t_1 in **diagram I**, the effect of automatic stabilisers helps dampen the fall in economic activity.

- When incomes fall and unemployment rises, falling marginal tax rates mean that tax burdens on households will decrease. This stimulates consumption.

- Lower AD and lower real GDP is associated with increased unemployment. Hence unemployment and

social benefits will also increase in the economy. This has a positive effect on households' spending which helps to limit the decrease in aggregate demand. This can help lessen the severity of a downturn/recession.

This is shown by the upward readjustment from AD_1 to AD_2 in **Diagram III**. (Remember; unemployment benefits and transfer payments are NOT included as elements of GDP, as this would cause double counting. However, when these transfer payments

help to *bolster consumption* and thus aggregate demand, there is a positive effect on GDP.)

The effect of automatic stabilisers over time is that the swings in economic activity, the amplitude of real GDP cycles, are somewhat milder than without, shown by the blue cycle in Figure 57.4 I. Note that automatic fiscal stabilisers do not solve recessions or overheating but merely lessen the impact of them somewhat.
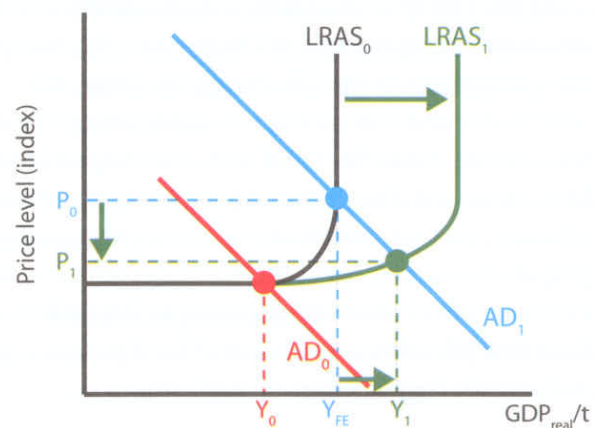
However ...

As mentioned in conjunction with *Effects of fiscal policies – Keynesian model* (see Figure 57.3) the degree to which government might influence aggregate supply depends on where initial macro equilibrium has been established. In Figure 57.5 there are two possible macro equilibriums; Y_0 and Y_{FE}. Posit that government spends money on building new roads, high-speed telecommunications networks and R&D facilities linked to universities. This would increase LRAS (**LRAS$_0$** to **LRAS$_1$**) but the outcomes are quite different for the two scenarios.

- If equilibrium is initially at Y_0, the increase in LRAS has no effect on either GDP, inflation or unemployment.

- At the full employment level of output (Y_{FE}) however, the shift in LRAS causes economic growth and benign deflation.

The conclusion is fully in line with standard Keynesian theory; at low levels of income, policies should be aimed at the demand-side rather than the supply side.

> ### Definition: 'Recession – revisited and criticised'
> A goodly number of my colleagues adamantly (= stubbornly) claim that the textbook definition of a recession ('... two falling quarters of real GDP ...') is not only unrealistic but highly out of date. There is merit in this view. The NBER (National Bureau of Economic Research – a very powerful US non-profit organisation where, amongst 16 other Nobel Laureates, Milton Friedman has submitted research) uses key monthly indicators of economic output, including employment, industrial production, real personal income, and wholesale and retail sales - to determine when economic growth has turned negative, rather than relying solely on two quarterly declines in GDP.

Fiscal policy and LR growth (AS... interventionist S-side policies)

We have dealt with government interventionist supply-side policies in Chapter 49, where improvements or increased quantities of natural, physical and human factors shifted the LRAS curve to the right (See Figures 49.2 and 49.3.) More specifically for fiscal policies and the supply-side, government can serve to create *pre-conditions for growth* by way of:

- Evening-out the business cycle. The increased predictability in the LR business cycle incentivises investment and FDI. The same goes for increased price stability.

- Government spending on education, health care, infrastructure enhances long run productivity.

- ... other supply-side policies of interventionist nature (see Chapter 61)

AD$_0$: LRAS shifts from LRAS$_0$ to LRAS$_1$ due to an increase in infrastructure and education yet equilibrium output at Y_0 means there is no effect on either GDP or unemployment.

AD$_1$: LRAS shifts from LRAS$_0$ to LRAS$_1$ causing an increase in quantity of aggregated demand – a new macro equilibrium is established at P$_0$ and Y$_1$. (Note that there is a negative output gap.)

Figure 57.5 Shifting LRAS in the Keynesian model

Evaluation of fiscal policy

"Predictions are rather hard to make – especially about the future." An observation credited to Niels Bohr.

Advantages of demand-side policies

The theoretical benefits of demand management have pretty much been outlined above, so consider this just a little more depth.

Macro goals

By adjusting monetary and fiscal policies, it is possible to influence the level of economic activity and therefore output, unemployment, inflation and the trade balance. In other words, demand management gives politicians tools to achieve the macroeconomic goals of society. In addition to this, the built-in fiscal stabilisers – *automatic* stabilisers – help even out the economic cycles and create stability and predictability in the economy, while evening-out some of the excesses/surpluses in productive capacity over time.

Target sectors of the economy

Fiscal policies can target specific industries, regions and labour groups – and often all three since there is a considerable overlap (revise structural unemployment in Chapter 51). For example, the labour government in the UK during the 1960s and '70s used numerous demand-side measures to lower unemployment levels. Nationalised industries (steel and coal) helped bolster demand for labour as did the increasing size of the government provided health care (National Health Service). Other methods (look up Japan for this one!) have often been large infrastructure projects to create jobs in under-developed areas. Keynes himself wrote about how job creation as a result of fiscal policies could be directed towards lagging urban or rural areas.[1]

Effects on AD

Keynesian economics also stresses the element of self-perpetuation that is stimulating aggregate demand via fiscal policy during recessionary periods; the **multiplier effect** (a HL concept, see Chapter 47). The effect is built into the Keynesian model and shows how the net final effect of increased government spending or lower taxes is increased over successive 'rounds' in the economy. For example, if government spending increases by £10 billion – say to build roads and other infrastructure – then jobs will be created and unemployment will fall. It doesn't stop there; many of those who have been unemployed for

some time will have pent-up consumption demands, so they will spend most of their wages. This spending will in turn create more demand, which creates more jobs, which lowers unemployment … etc. Thus, according to Keynesians, a major benefit of demand-side economics is that there is a 'leverage' effect of using fiscal stimulation, since the final increase in national income is greater than initial budget costs; $\Delta\uparrow Y_{final} > \Delta\uparrow G_{initial}$ (or $\Delta\uparrow Y_{final} > \Delta\downarrow T_{initial}$). The increase in national income will ultimately pad government tax coffers and help to make up for possible deficit spending done in the first place.

Promoting economic activity during recession

Discretionary fiscal policies in turn allow governments to steer the economy in line with both consensus views of economic goals and ideological underpinnings and ideals of social/economic welfare. Full employment increases living standards while tax rates, unemployment benefits and government spending all help in improving social welfare systems and redistributing incomes in the economy. Note that the concepts such as 'fairness' and 'equality' are NOT entirely normative. There are in fact a good many sound economic arguments underlining social redistribution by way of taxes and transfer payments, in that a great many economic and social costs show strong positive correlation to increased inequality of income and wealth; crime, alcoholism and drug abuse are notable examples.

Disadvantages of demand-side policies

Keynesian demand management was virtually unchallenged as the macroeconomic method of choice during the 1950s and '60s. High growth rates and low unemployment seemed to justify demand management policies. However, during the early 1970s, a number of weaknesses of demand management became evident.

- The business cycles were often erratic and 'untamed' or even aggravated by demand management policies – there were increasing indications of *politically* inclined business cycles over the course of changes in governments.

- *Inflation* rose to hitherto unseen levels and budget deficits grew since government spending during recessionary periods was evermore seldom made up for during booms.

---

[1] *Fiscal Policy: Why Aggregate Demand Management Fails and What to Do about It*; Pavlina R. Tcherneva, Levy Economics Institute of Bard College, January 2011 (Working paper no 650)

- This increased government *indebtedness* led to most debilitating exchange rate problems (see Chapter 67). Increasingly open economies meant that government spending and/or lower taxes would not have the same multiplicative effect on the domestic economy, as an increased proportion of disposable income gains flowed out of the country to buy imports.

These main weaknesses in using demand management to control the economy are given below under five headings: Trade-off problems; time lags; new-classical critique; political constraints; and limits of fiscal policies in dealing with supply-side shocks.

Trade-off problems

The purpose of fiscal (and monetary) policies is of course to control the swings in the business cycle and achieve macro objectives. This has been illustrated earlier but let's now look at the possible opportunity costs – *trade-offs* – that arise as a result.

- **Growth and inflation:** High growth and low unemployment are key macroeconomic goals. Governments can use expansionary fiscal policy to stimulate the economy where there is below full employment. See *Case study: Fiscal policy in the US* earlier. The aim of these tax measures is to stimulate consumer spending and investment spending and increase aggregate demand. This creates inflationary pressure which results in higher final prices and a *trade-off* between growth and inflation in the short run. (Note that price stability does not imply zero inflation but low and steady inflation.)

- **Budget and unemployment:** When unemployment rises the government can decrease taxes and/or increase government spending in order to increase aggregate demand. Since stimulatory fiscal measures are often undertaken during recessions, they often result in *deficit spending*, resulting in a trade-off between a balanced budget and unemployment levels.

- **Growth and trade balance:** In stimulating the economy, incomes increase and for reasons explained earlier imports will also increase due to the propensity of citizens to import goods and services. An economy experiencing a *trade deficit* (or, more correctly, a *current account deficit*; see Section 4.5) might use fiscal policy to adjust the trade imbalance. Since

high growth rates are associated with both increased inflation and increased imports, tighter fiscal policy could be instigated to counter both inflation and trade imbalance. The contractionary policies would thus lower growth in order to lower import expenditure. A government will have difficulty in keeping the trade balance in equilibrium while stimulating domestic growth and employment – something that the US has noticed for all but *one* of the past 22 years.

- **Interest rates and exchange rate:** Finally, there is both a domestic and foreign sector trade-off arising from monetary policy. When the Central Bank decreases interest rates in order to stimulate the economy, there will be downward pressure on the *exchange rate* as foreign investors/speculators pull some of their funds out of the country in order to place them elsewhere at a relatively better rate of interest. This will lower demand for the currency and thus the price of the currency falls – which is nothing other than the exchange rate. The Central Bank therefore cannot set the deomestic currency's value to other currencies and still freely use monetary policy as a demand-management tool. This was a main argument for European Union opponents of the single currency, the EURO, since monetary policy would in effect be taken out of domestic hands and surrendered to the European Central Bank (ECB) in Frankfurt.[2]

Five possible macroeconomic trade-offs emerge from the discussion above:

A. Growth ⇔ price stability

B. Unemployment ⇔ price stability

C. Unemployment ⇔ balanced budget

D. Growth ⇔ trade balance

E. Domestic monetary policy (interest rate) freedom ⇔ stable (or fixed) exchange rate

Time lags and exacerbation of the business cycle

In trying to attain these goals, governments can actually make things worse by exacerbating (= worsening, intensifying) the business cycle. It is possible that attempts to fine-tune

2   Frankfurt **am Main** – not the other Frankfurt.

the economy result in 'stop-go' cycles, where the economy frequently moves between boom and recession. New-classical economists are quick to point out that *fiscal stabilisation policies* often come as 'pouring gasoline on a growing fire and water on dying embers' – i.e. that foreseeing correct demand management policy is virtually impossible due to time lags.[3] There are a number of possible lags built into the macroeconomic system, all of which make it most difficult to steer – 'fine tune' – the economy.

Identification lags arise simply because it is always difficult to see where you are on the business cycle, i.e. the depth of a boom/recession and the length. In Figure 57.6, the recognition stage when contractionary policies should be instigated is at point I.

It will take time for the political and administrative process to result in actual policy decisions, leading to *decision and implementation lags*. Say that the government decides to tighten fiscal policy to deflate the overheating economy at point II …

… it will still take time before the effects of higher taxes and/or lower government spending actually has an impact on the economy. *Effect and impact lags* are possibly the most heavily weighted of all three, since the resulting change in aggregate demand might deflate an already contracting economy. This is shown by point III in the diagram, where tighter fiscal policies kick in and lower real GDP at a faster rate than would otherwise be the case.
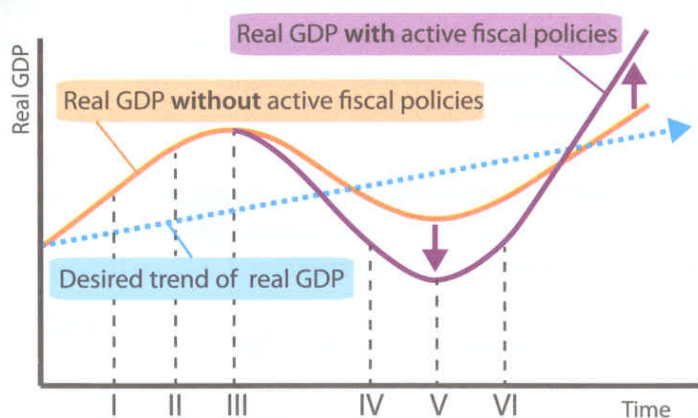

Figure 57.6 How lags can exacerbate fluctuations in the business cycle

3 A famous allegory of economists' attempts to predict/control the business cycle is that it's like trying to drive a car where all the windows are blacked out and you try to steer by looking in the rear view mirror. In addition to this, the accelerator (throttle) and brakes are very sluggish and therefore slow to react, leading to time lags between putting your foot on them and the reaction of the car. Finally, the car has exceptionally good shock absorbers so that you cannot feel whether you are going up or down. Now try to keep on the road. No, don't try this at home kids.

The exaggerated cyclical effect resulting from contractionary policies in Figure 3.4.14 continues in the next stage of recession. The identification lag (IV) is followed by an implementation lag (V) and an inflationary policy that takes effect at point VI, where the economy has already started to recover. The effects of these reflationary lags are like 'pouring gasoline' on a growing fire, once again causing GDP to increase more acutely than otherwise. The effect over an entire cycle is that real GDP fluctuates more than would otherwise be the case.

New-classical critique

This subject could fill an entire book on its own. Come to think of it, it has filled *hundreds* of books. However, I will limit the discussion here to three main points of criticism central to the supply-side reaction against demand management; *inflation* in the long run, market *inefficiency* and *crowding out*.

Inflation vs. long run growth

Classical economics points to the *stagflationary* (= rising inflation and low/falling GDP) demand-side period of the 1970s and the period of price stability and growth of the supply-side 1980s. Demand-side policies ultimately failed quite drastically in balancing budgets over a cycle, as deficit spending during recessions was never countered during boom periods. Demand-side policies therefore have their place in combating inflation, not in increasing output, according to the new-classical view. New-classical economists put heavy emphasis on using supply-side policies rather than demand-side policies to increase long run growth. (See Chapter 46, Figure 46.5; New-classical view … to see how demand-side policies are considered ultimately inflationary.) This viewpoint has to a certain degree become mainstream and most economists would agree today that there is merit in using an element of supply-side policies in order to lower unemployment in the long run.

Demand-side policies lessen market efficiency

Another basic premise of the new-classical school is the inter-temporal opportunity cost issue of demand management, where demand-side policies focusing on alleviating unemployment will have far greater long run costs when inflationary pressure not only dissolves any short run gains in income, but where resources are squandered (= wasted) on inefficient government spending. The interventionist leaning of demand management distorts factor markets and inhibits market clearing. The new-classical school prescribes the use of production incentives such as lowering corporate taxes and labour incentives such as

decreasing income taxes and lowering social/unemployment benefits. The short run social and economic costs of these policies would be more than compensated for by increased long run output and full employment levels. One could say that new-classical economists are inclined to let the golden eggs hatch and produce more geese rather than using the golden eggs.

Crowding out

There are basically three ways government can fund increased government spending during a fiscal year; printing money, raising taxes, and borrowing from its citizens. Printing money is relatively straightforward and most economists would agree that this indeed raises output at least in the short run. Most would also agree that raising taxes to fund spending has a relatively minor effect on aggregate demand. Instead, the disagreement has centred on whether government spending *funded by borrowing* will bring about an increase in aggregate demand or not.

The argument goes as follows: Assume that the government borrows money (by issuing government securities, i.e. bills and/or bonds) in order to fund government spending in line with demand-side fiscal policies. The increase in government borrowing increases the demand for loanable funds which drives up interest rates and causes investment expenditure to fall. Thus, the potential increase in aggregate demand due to government spending is negated – 'crowded out' – by the increase in interest rates and concomitant fall in investment;

hence the name **crowding out**. The concept has frequently been used as another new-classical/monetarist argument against fiscal policies.

> **Definition: 'Crowding out'**
> When government expenditure is financed by increased government borrowing interest rates may be driven up. This might cause a decrease in investment in the private sector as firms scale back on capital expenditure. The increase in increase in government expenditure and borrowing has 'crowded out' an amount of investment.

However (this is economics; there is *always* a 'however'), there are a couple of rather weighty **assumptions** involved in claiming that government borrowing would crowd out investment.

1. Firstly, the economy has to be operating **at or above the full employment** level of output in order for complete crowding out to take place.

2. Secondly, **real borrowing** has to take place; the government's increased borrowing from either households or the financial sector cannot be compensated for by pumping additional money into the market via the printing press.
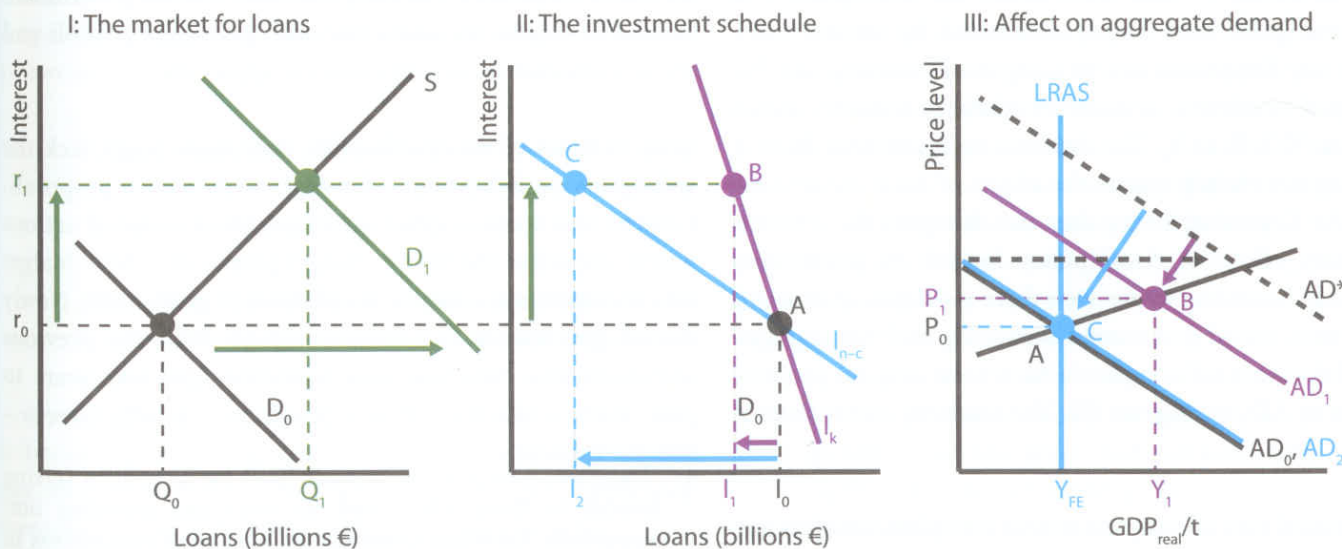
MACROECONOMICS

I: The market for loans II: The investment schedule III: Affect on aggregate demand

Figure 57.7 Crowding out – Keynesian and new-classical view

1. Assume that the economy is operating at the full employment level of output, Y_{FE} in **figure 57.7, diagram III**, and that government increases government spending by way of increasing its borrowing on the open market. The increase in government spending increases aggregate demand from AD_0 to AD^*, but …

2. … the increase in demand for loanable funds (**diagram I**) caused by government borrowing will drive up the interest rate from r_0 to r_1, which in turn decreases investment …

3. … shown in the two different investment schedules in **diagram II**. This has a *contractionary* effect on aggregate demand; AD_1 or AD_2 in **diagram III**. The movement from **A to B** or **A to C** in **diagrams II and III** are two of a number of possibilities.

Summing up in economic shorthand: $\Delta\uparrow G \rightarrow \Delta\uparrow D_{loanable\ funds} \rightarrow \Delta\uparrow r \rightarrow \Delta\downarrow I \rightarrow \Delta\downarrow AD$… the increase in government spending drives up interest rates which 'crowds out' investment. The question of crowding out is largely one of degree. Most economists would agree that there is some crowding out when government borrows money to fund additional spending, but there is a great deal of contention as to the extent to which investment funding is affected.

Keynesian view of partial crowding out: Keynesians traditionally largely disregarded any possible crowding out effects, operating as they were under the assumption that investment levels were largely unaffected by interest rates. Modern-day Keynesians accept a *degree* of crowding out, but still regard *investment demand as relatively inelastic*, shown in **Figure 57.7 II** as I_K. The decrease in investment from I_0 to I_1 does not entirely negate the effects of fiscal stimulation. Hence the Keynesian theory does not disregard the potential stimulatory effects of fiscal policy funded by government borrowing. According to this view there is a *degree of crowding out* due to increased government borrowing, and that aggregate demand does not fall completely back to its original position, AD_0, but to AD_1 in **diagram III**. The economy moves from **A to B**.

New-classical view of complete or near-complete crowding out: This view claims that complete – or near-complete – crowding out will indeed occur (as long as there is *real borrowing* taking place) since the demand for investment is *highly responsive* to changes in interest rates. The *elastic investment schedule* of the new-classical/ monetarist school, I_{n-c} in **Figure 57.7 II**, causes a far greater decrease in private sector investment, from I_0 to I_2.

In the example here, complete crowding out takes place. The results are well in line with the new-classical view that fiscal policy squanders resources by diverting funds from the private sector to government.[4] The macro outcome is that the increase in government spending aims to increase aggregate demand from AD_0 to AD^*, but since there has been *real borrowing* in order to fund the increase in government spending, the increase in the interest rate concomitantly decreases investment – shown by the shift of aggregate demand from AD^* back to AD_2. The stimulatory effect of government spending is thus entirely eradicated by a decrease in private sector borrowing and thus lower investment levels. There is *complete crowding out* and the economy moves from **A to C**.

Political constraints

"Government's view of the economy could be summed up in a few short phrases: If it moves, tax it. If it keeps moving, regulate it. And if it stops moving, subsidise it." Ronald Reagan

Anyone who saw some TV news during 2011 probably saw the demonstrations in Greece, where thousands of citizens converged at Syntagma Square in Athens. That's what happens when government cuts spending and lays off thousands of public employees because debt is over 150% of GDP, the public sector employs close to 25% of the labour force and spends 80% of public monies on wages, pensions and social security benefits for public employees. The facts indicated that the government needed far harsher measures than were politically possible and this is a situation most politicians can appreciate.

Even in times of reasonable calm, politicians might lack the willingness or ability to put forward much-needed proposals. Limited time spans in office can create 'short-termism' actions which jeopardise the future. Budget proposals – both budget cuts and increases – can be voted down in parliament. Newly elected government can 'inherit' deficits from the previous administration. New laws and regulations can take years to pass. There is also the very real possibility – as with Greece –

---

4   There seems to be a link between 'pork barrel politics' (giving favours to those who voted for you) and crowding out! Apparently American congressmen who chair committees in congress make sure that the home state gets a lion's share of any federal funds – some 40 – 50% more than without a home-state congressman. This has had a negative effect on private firms in the region which reduce capital expenditure by 8 – 15%. Increased government spending has crowded out a degree of private sector investment. (Source: Economist, Oct 30 2010, *Far from the meddling crowd*)

that membership in a Free Trade Area or Customs Union limits the degree of manoeuvrability for a government wishing to borrow for increased government spending – this is a theme for Chapter 74.

Who took my pension plan?! (Syntagma Square, Athens, June 2011 GNU license)

Limits of fiscal policy – S-side shocks cannot really be dealt with

The limitations of demand-side policies became notoriously clear during the stagflationary period after the first 'oil shock' in 1973/74. Thirty years of reasonably successful demand management left little preparedness for a situation that meant rising unemployment *and* rising inflation. (Revise Chapter 53.) Any fiscal demand-side policies means choosing the lesser of two evils:

- Government can increase spending and/or lower taxes to stimulate growth – but this means increasing inflation which is already rising.

- The other possibility is to dampen inflation by contractionary fiscal policies – and of course this worsens growth and unemployment.

It became clear during the 1970s that governments should pay increased attention to long run effects and supply-side measures. These will be dealt with in Chapters 60 to 62.

Summary & revision

1. Expansionary fiscal policies include lowering taxes (for example income and profit taxes) and increasing government expenditure – both increase AD. Contractionary fiscal policies thus decrease AD.

2. Fiscal policies are discretionary (decisions made by government) and automatic (marginal tax effects and transfer payments) which are built into the system.

3. Government can decrease inflationary gaps using contractionary policies and close a deflationary gap using expansionary policies.

4. The effects of fiscal policies on real GDP and inflation vary along the Keynesian AS curve:

   a. At low levels of income and high unemployment, an increase in AD might increase GDP without inflation

   b. Nearing the full employment rate of output, AD stimulation renders a trade-off between inflation and unemployment

   c. At the full employment level of output any fiscal expansionary policies will be purely inflationary

5. Modern economies have so-called automatic stabilisers built in to the system. Increasing AD and higher incomes automatically lower social and unemployment benefits and increase taxes on income and profits – all of which have a dampening effect on AD. Conversely, falling AD means higher unemployment and social/unemployment benefits kick in automatically while income and profit taxes fall – this serves to stimulate AD during downturns in the economy.

6. Fiscal policies can also affect AS by increasing the stability of the economy to incentivise investment. Spending on education and infrastructure are also supply-side (interventionist) policies.

7. The Keynesian AS-AD model shows that stimulating **LRAS might have zero or limited effect at very low levels of income** – e.g. if macro equilibrium is along the horizontal portion of the Keynesian LRAS curve.

MACROECONOMICS

8.   Advantages and **strengths of demand-side** policies:

   a.   Government can to some extent achieve the **main macro goals**

   b.   Sectors and regions of the economy can be **targeted**

   c.   There is a possible **multiplicative effect** of increased government spending (HL)

   d.   Both discretionary and automatic fiscal policies can help **limit recessions and periods of overheating**

9.   Disadvantages and **weaknesses of demand-side** policies:

   a.   The 1970s taught governments that demand-side policies often result in *inflation, stop-go cycles* and *deficits/debt*

   b.   Numerous **trade-offs** exist in implementing demand side policies, the most obvious ones being growth ⇔ inflation and unemployment ⇔ inflation

   c.   **Time lags** can actually exacerbate business cycle fluctuations

   d.   **Political constraints** often limit the degree and speed of implementing much-needed fiscal austerity measures

   e.   Fiscal policies are inadequate in dealing with **supply-side shocks**

   f.   The **new-classical school** of economics criticises demand management on several counts; long run *inflation* and resulting *lack of competitiveness, decreased market efficiency* and …

10.   …. **crowding out**. A key new-classical/monetarist argument proposing that any increase in real borrowing by government to fuel AD will drive up interest rates and 'crowd out' an amount of private sector investment. This will of course lessen the degree to which the demand-side stimulation works.

58. The Central Bank and Monetary Policy

2.5

Key concepts:

- Role of central banks
    - Regulator of commercial banks
    - Lender of last resort
    - Government's bank
- Central bank responsibilities
    - Interest rate
        - Discount rate
        - Supply of money
    - Exchange rate policies

All money is a matter of belief.
Adam Smith

Role of central banks

There are a number of functions of the Central Bank, whereof four are central to the monetary policy dealt with here: as a regulator of commercial banks, acting as the 'banks' bank', functioning as the lender of 'last resort' and acting as the government's bank.

Regulator of commercial banks

Central banks are charged with **regulating commercial bank lending**, an example of which is setting minimum levels of cash which commercial banks must keep on hand, so called 'minimum reserve requirements'. There is also a *judicial*

function since central banks will oversee financial regulations concerning commercial banks' lending practices and accounting methods.

Banks' bank

One of the most important tasks for central banks is to be the 'banks' bank', i.e. seeing to it that commercial banks' need for cash is met. This is done at an interest rate known as the **discount rate**[1] which is a key in setting market interest rates. (See *Central Bank and interest rates* and figure 58.1 following.) In many countries, commercial banks are *required* to keep a certain amount of cash deposited with the central bank.

Lender of last resort

The central bank is in fact where the buck stops.[2] When financial institutions are in financial difficulties and simply cannot obtain loans elsewhere, the central bank can step in as the 'last resort'. The reasoning is pretty straightforward; a collapse of several large financial institutions will have serious repercussions on the economy and will ultimately hit households the hardest – as was the case when almost half of the banks went bankrupt during the Great Depression in 1930s USA. The central bank

---

1   There is a most confusing array of different terms and forms of interest that central banks charge commercial banks for loans. I will simply use 'discount rate'.
2   I am referring to a famous plaque on US President Truman's desk. "Passing the buck" means to shift responsibility on to somebody else. Truman had a plaque made saying "The buck stops here!"

basically aims to provide stability in the economy by avoiding mass panic withdrawals when banks become illiquid. You can just imagine how commercial banks and financial institutions avoid this last resort – the interest rates are at penalty level and it doesn't do wonders for the bank's reputation.

Government's bank

Perhaps the most visible role of central banks in terms of media coverage is managing government deficits and debt. When the government needs to finance deficit spending it issues government bills and bonds (IOUs) through the central bank. These are then purchased by financial institutions and to a certain extent households.

Central bank responsibilities

In addition to the above, central banks are responsible for issuing note and coins ('printing money'), managing the foreign currency reserves, servicing government debt, and managing monetary policy. We will focus on the issue of *monetary policy* and how it is used to achieve the main macro objectives of price stability (low and stable inflation), growth, employment and exchange rate stability.[3]

---

3   Note that the core objectives for central banks vary between countries. Most will have a focus on maintaining low inflation and low unemployment.

Interest rate

Central banks implement **monetary policy** through goals set by government which are often in the form of inflation targets. This makes the relationship between interest and inflation somewhat tricky since cause and effect works both ways; an increase in inflation causes interest rates to rise, yet an increase in the interest rate is used as a monetary tool in order to reduce inflation. The interest rates in an economy are set by government using two methods; the *discount rate* and the *supply of money*.

Discount rate

Figure 58.1 illustrates how a Central Bank can influence market interest rates. Let's follow a simple lending flow by assuming that there is only one commercial bank which is lending to households and firms at an interest rate (the 'price of money') of 7%. Now, commercial banks make a profit by having as much of their money assets in turnover as possible, i.e. they want to lend as much money as possible. In order to have loanable funds available, commercial banks offer households and firms an incentive; a return on deposited money, i.e. an interest rate. In the example below, the interest rate received is 2%. The banks basically make their money by borrowing money from you and your business at 2% interest and then lending it to your neighbour and her business for 7% - the 5 percentage points being the banks' profit margin.
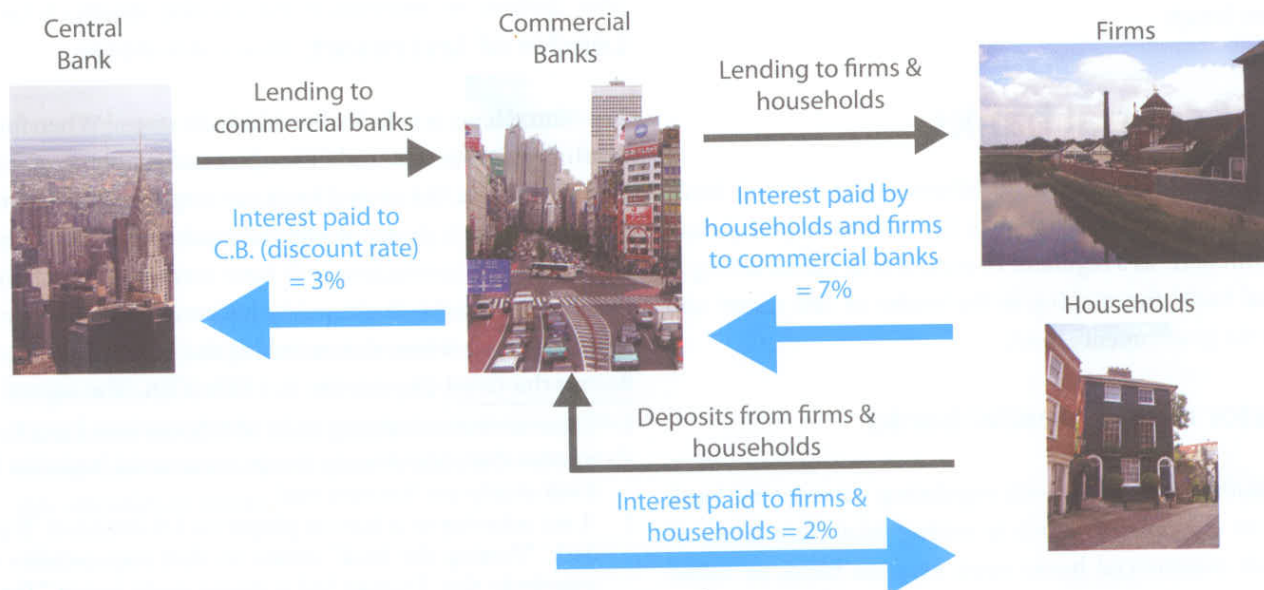

Figure 58.1 Central bank and interest rates

If the central bank raises the interest rate at which commercial banks borrow, the discount rate, the profit margins of commercial banks will fall. The commercial banks will therefore raise the lending rate for firms and households borrowing money. The deposit rate at commercial banks – what households and firms receive for depositing money – will also rise.

Supply and demand for money

The **demand for money** does not show 'how much money we want'. It shows the propensity of firms and households to want to hold their assets in *liquid* form (e.g. money) rather than in non-monetary forms such as machinery, cars, shares, property ... or, in my case, wrist watches. The 'price' of money is the rate of interest and the demand curve for money (also known as the 'liquidity preference schedule') is downward sloping, signifying that the lower the interest rate the higher the quantity of money will be demanded.

The narrow definition of the money supply is the total amount of notes and coins circulating in the economy plus the balances held by commercial banks. The supply of money is assumed to be vertical (here) in that it is exogenous – created outside the system and set by monetary policy. The complete interest inelasticity of the money supply is based on the simplifying (and rather unrealistic) assumption that the supply of money is totally controlled by the Central Bank's monetary policy. Figure 58.2 shows how the supply of money (**Sm**) and demand for money (**Dm**) create an equilibrium interest rate of r_0 and **Qm**.
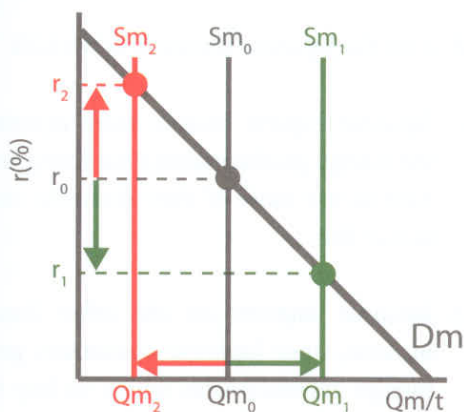
Δ↓Sm: An increase in the discount rate, the selling of government securities or an increase in the reserve requirement of commercial banks decreases the supply of money (Sm_0 to Sm_2)

Δ↑Sm: A decrease in the discount rate, the buying of government securities or a decrease in the reserve requirement of commercial banks increases the supply of money (Sm_0 to Sm_1)

Figure 58.2 Supply and demand for money

Changing the supply of money: While interest rates are the main instrument in monetary policy, Central Banks can use several other instruments to influence the supply of money and thereby the rate of interest. The Central Bank commonly has a monopoly on printing money, and thereby total control over the amount of outstanding notes and coins, i.e. it is possible to 'print money' (which the Central Bank technically then lends to the government) in order to fund government spending. The three most common methods for controlling the supply of money are *discount rate* changes, *open market operations*, and *direct controls* on lending.

- **Discount rate:** As described earlier, the discount rate is the Central Bank's rate of lending to commercial banks. An increase in the **discount rate** reduces the willingness of commercial banks to borrow from the Central Bank and thus limits their ability to lend money, which serves to lower the amount of loans. This decreases the money supply (Sm_0 to Sm_2 in *figure 58.2*) and drives up interest rates (r_0 to r_2).

- **Open market operations:** The Central Bank can conduct **open market operations** on the financial market much like any other financial institution; it can issue government debt, 'IOUs'[4], in the form of interest-bearing bills and bonds (collectively known as government securities). When the government sells securities, they can be purchased by both the financial and non-financial sector (firms, households and other governments). When these securities are purchased, the money supply is decreased since funds in terms of cash are 'soaked up' from the market (Sm_0 to Sm_2) – the rate of interest increases from r_0 to r_2. If instead the Central Bank buys (back) securities on the open market, there is an increase in the supply of money (Sm_0 to Sm_1) and a fall in the rate of interest (r_0 to r_1).

- **Direct controls on banks' lending:** In its capacity of regulatory body for the financial sector, the Central Bank sets certain limits on the amount of outstanding loans (= total loans given to households and firms) banks are allowed to have, some form of **reserve requirement**. A reserve requirement of 5% means that a bank which has maximised its lending capacity and has €100 million in deposits must keep €5 million in

4   An IOU is a layman's (= normal person's) term. It is a bill of debt issued by the lender (creditor) and signed by the borrower (debtor). The debtor now owes money and the creditor has a piece of paper verifying this which basically says; "I owe you" = "IOU".

cash reserves. If the Central Bank wishes to decrease the supply of money, the reserve requirement can be increased; say to 6%, whereby the bank in the example will have to limit new loans until it once again has the minimum requirement of cash on hand. An increase in reserve requirements will decrease the supply of money (Sm_0 to Sm_2) and increase interest rates from r_0 to r_2. A decrease in reserve requirements will enable banks to increase lending, which increases the money supply and lowers the rate of interest.

Final notes on money supply and interest: The causal relationship between the money supply and interest rates is *extremely* complicated. It is impossible to set a precise supply of money and very difficult to foresee the extent to which interest rates will change as a result of adjusting money supply. Also, while basic economic theory stipulates that a change in the supply of money causes a change in interest rates, there is heavy empirical evidence of a stronger reverse causal flow, i.e. that a change in interest rates causes a change in the supply of money. As a point of fact, central banks often use the rate of interest to adjust the supply of money rather than the reverse. While this is clearly outside the boundaries of your syllabus, it is well worth noting that Central Banks can set the money supply (and let interest rates adapt) or set interest rates (and allow the supply of money to adapt) – but it cannot set BOTH the rate of interest AND the supply of money.

You may have noticed how I have used the plural form mostly: 'interest rates'. This is not sloppiness on my part, as there are a good many different interest rates. Mortgage rates for housing; short-term rates from banks; government rates on bills, bonds; repo rates; discount rates … I have a list of some 50 different interest-bearing instruments – all of which will be *interdependent*. You see, while interest rates will most definitely vary (primarily short- and long-term rates), there will be very noticeable *co-movement in all the different market rates* offered. If the Central Bank raises one of its interest rate forms, then there will be concomitant (= associated) upward movement amongst all other interest forms on the market. I bring this up in order that you do not get confused when you are confronted by the myriad (= multitude) of different rates flickering by in the media.

Exchange rate policies

Although it is a main theme of Chapters 67 to 69, I cannot help but briefly bring up the subject of how monetary policy can affect a country's exchange rate and thus the domestic economy. Consider an economy operating in equilibrium beyond the full

employment level of output, i.e. where there is an *inflationary gap*. The Central Bank can use monetary policy as outlined above to increase the interest rate and lower consumption and investment in the economy, which exerts downward pressure on aggregate demand. There will be an additional dampening effect on aggregate demand arising from the effect on the exchange rate of the domestic currency. In other words, increased interest rates and the international substitution effect (see Chapter 40) will help to create additional disinflationary pressure on the domestic economy.

Say the Central Bank of Japan (Bank of Japan, BOJ) raises interest rates. Foreign firms, financial institutions, investors/speculators and governments will see that by readjusting a portion of their foreign currency holdings in their international portfolio of assets (which is comprised of foreign assets such as currencies, shares and securities) and buying Yen (¥), an increased rate of return is possible. This, naturally, assumes *ceteris paribus* in that all other countries' interest rates are left unchanged.

- The *demand for Yen will rise*, since foreigners will have to first purchase Yen to deposit in Japanese banks.

- This increases demand for the Yen which *drives up the value of the Yen*. (The price of the Yen is put in terms of other currencies, say the US dollar.) Thus the Yen increases in value vis-à-vis other currencies, it **appreciates** (increases in price in terms of the US dollar).

- The effect on the Japanese economy is twofold:

  o Japanese *exports become more expensive* on the foreign goods market since foreigners now have to use more of their domestic currency to buy Yen.

  o Japanese *imports on the other hand will increase*, since Japanese consumers get more foreign currency with which to buy foreign goods, they will substitute some domestic goods with foreign.

Decreased export revenue and increased import expenditure in Japan will cause aggregate demand to decrease and have a disinflationary effect on the Japanese economy.

The mechanisms linking monetary policy to the exchange rate and exports/imports are frequently employed by central banks to adjust aggregate demand. Since a decrease in interest

rates puts downward pressure on the exchange rate, export dependent countries have often used interest rates to decrease the exchange rate and increase demand for their export goods. While this may sound relatively straightforward, there are a number of negative effects arising in an economy experiencing a falling exchange rate. These are looked at in Chapters 67 to 69.

Summary & revision

1. The role of central banks is to *regulate commercial banks*, be the commercial *banks' bank*, act as *lender of last resort* and be the *government's bank*.

2. Central banks' responsibilities are primarily to **implement monetary policy** aimed at achieving macro goals such as price stability, growth, low unemployment and exchange rate stability.

3. The main tools of monetary policy are setting the **interest rate** and the **supply of money**.

4. Interest rates are influenced primarily by setting the rate at which commercial banks borrow money (the **discount rate**) and the **money supply**.

5. Central banks can **adjust the supply of money** in three ways:

   a. Changing the *discount rate*

   b. *Open market operations* (buying or selling government securities)

   c. Changing commercial banks' *reserve requirements*.

6. Another key objective – that of **stability in exchange rates** – can be attained by central banks by adjusting the interest rate:

   a. An *increase in the interest rate* attracts foreign deposits and thus increases the demand for the Home currency and the *Home currency appreciates* (rises in value compared to other currencies)

   b. A *decrease in interest rates* causes an outflow of Home currency to foreign banks where the interest is relatively higher – the supply of the Home currency increases causing a *depreciation*.

7. A **key trade-off in monetary policy** is that it is impossible to freely set the interest rate while simultaneously setting the exchange rate.

MACROECONOMICS

59. The Role of Monetary Policy

Key concepts:

- Interest rates and aggregate demand
    - Expansionary (loose) and contractionary (tight) monetary policy
- Inflation targeting
- Evaluation of monetary policy
    - Independence of the central bank
    - Time lags
    - Keynesian model
    - Limits of monetary policy
        - Liquidity trap
        - Deep recession
    - Policy trade-offs

If you would know the value of money, go and try to borrow some. Benjamin Franklin

Interest rates and aggregate demand

If you lend someone $100 and get it back in a year, has it cost you anything? "Inflation would mean that the real value of the $100 is less" you might reply. OK, but what if we assume that inflation during the year was zero? Of course you would still incur a cost; the *opportunity cost* of not having access to the $100. You make up for this opportunity cost by charging interest.[1]

Definition: 'Real interest rate'

If nominal interest is eaten up by inflation then there has been an opportunity loss for the lender. It is therefore more relevant, once again, to focus on real values rather than nominal.

The **real interest rate** is the nominal rate minus inflation; $r_{real} = r_{nominal} - $ **inflation.**

Real interest

Assuming that your next-best option is estimated at a money value of $5, then you would charge an interest rate of (at least) 5% ($100 · 5% = $5). If we now assume that there indeed is a steady rate of inflation, say 5%, what would you do? Correct; you would charge a *higher* rate of interest, 10.25%[2], since receiving $105 in a year's time when the price level has gone from 100 (indexed at the time of the loan) to 105 means that the $105 you get back is worth only $100 in real terms.

The above illustrates a causal flow from inflation to interest rates; when inflation rises, commercial banks will raise interest rates to keep a given rate of *real* interest. However, when the central bank uses monetary policy to increase interest rates (**tight monetary policy**), the result is higher costs for borrowing and higher opportunity costs (foregone interest) of consumption. Higher interest therefore lowers consumption and investment and hence decreases aggregate demand.

When inflationary pressure rises, the central bank can implement **loose monetary policy** by lowering interest rates or decreasing the supply of money (which increases interest rates). Investment becomes cheaper for firms and households can increase their borrowing for consumption goods – aggregate demand increases. There is also the issue of lower opportunity costs of consumption since savings do not generate the same interest as earlier – this too incentivises increased consumption.

---

1   During an 'Open House Day' at school, I gave a lecture on economics to the parents of my people. I brought up the issue of opportunity cost and interest. The next day one of my students, Jakob, came back with a bottle of Polish vodka (Zubrowka, my favourite) as a thank-you gift from his mother. He wasn't too pleased about things; "Thanks a lot, Matt. My mom's now charging me interest on money she lent me to buy a computer."

2   Since you want to have $105 in **real** terms, you must charge an interest rate that gives you $110.25, i.e. 10.25%.

OUTSIDE THE BOX

The relationship between interest and investment

While it is not *directly* part of the syllabus, a most useful theoretical concept in understanding underlying forces connecting interest rates with aggregate demand (and also aggregate supply as shall be seen) is the downward sloping demand for investment, called the **investment schedule** (= demand curve for investment). Lower interest rates induce firms to increase investment for two reasons.

1. **Opportunity cost issue:** When the interest rate falls, a number of investment opportunities previously considered unprofitable are suddenly profitable. For example, assume that firms in an economy have only two options for their retained profit (= profit held over from previous years): put it in the bank or invest it in the firm. Now, say that firms in an economy itemise and assess all possible investment opportunities and that the current market interest rate is 7%; any investment which does not yield a rate of return of 7% or more will be put in the bank, since an investment yielding, say, 6.5% will render the firm an opportunity cost (loss) of 0.5%. Figure 59.1 shows how total investment demand during the year is €10 billion at an interest rate of 7%.

Say that the economy slows down and the Central bank loosens monetary policy and interest rates fall (as outlined earlier) from 7% to 6%. Investment options that were previously considered unprofitable suddenly become more attractive; what firm would leave money in the bank at 6.5% interest when an investment opportunity yields 6.9% - or even 6.55%?! Firms will therefore increase investment when the opportunity cost of investment (interest being the alternative) falls.
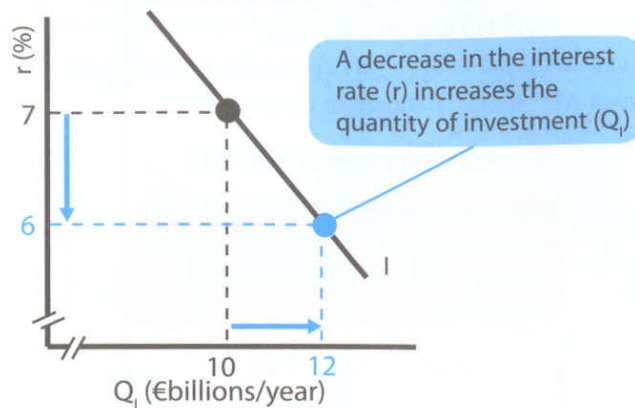
Figure 59.1 Interest rates and quantity of investment (investment schedule)

2. **Cost of investment:** Not all investment is funded internally, i.e. by held-over profits accumulated within the firm. A great deal of the funding for investment comes from banks and other financial institutions in the form of loans. When the interest rate falls, the cost of servicing debt (paying interest on loans) goes down and firms will be willing to take on more debt in order to invest.

The lower rate of interest in this example causes total investment in the economy to increase from €10 billion per year to €12 billion per year. The inverse relationship between interest and quantity of investment is thus a downward sloping curve, the investment schedule.

Expansionary (loose) and contractionary (tight) monetary policy

When unemployment rises, growth falters and/or inflation shows signs of falling below the floor of a targeted interest corridor due to decreasing aggregate demand, the central bank can implement **loose monetary policy**[3] to stimulate the economy. The central bank decreases the discount rate from 3% to, say, 2.5%. Commercial banks now have cheaper credit on money borrowed from the central bank, and will compete with each other by lowering the lending rates for firms and households. If commercial bank rates fall to 6.5% there will be a lower

3 Many of my non-native English speakers struggle with the words "loose" (= slack, movable) and "lose" (= drop, misplace) – all too frequently mixing them up. One of my colleagues at the Oxford Study Courses has experienced the same thing and came up with "I keep *loose* change in my pocket so I don't *lose* it."

MACROECONOMICS

opportunity cost associated with investment and consumption – so savings will fall and investment and consumption will rise. This causes *the fall in aggregate demand to decelerate* (the 'hook' between AD^* to AD_1), shown in diagram 59.2.
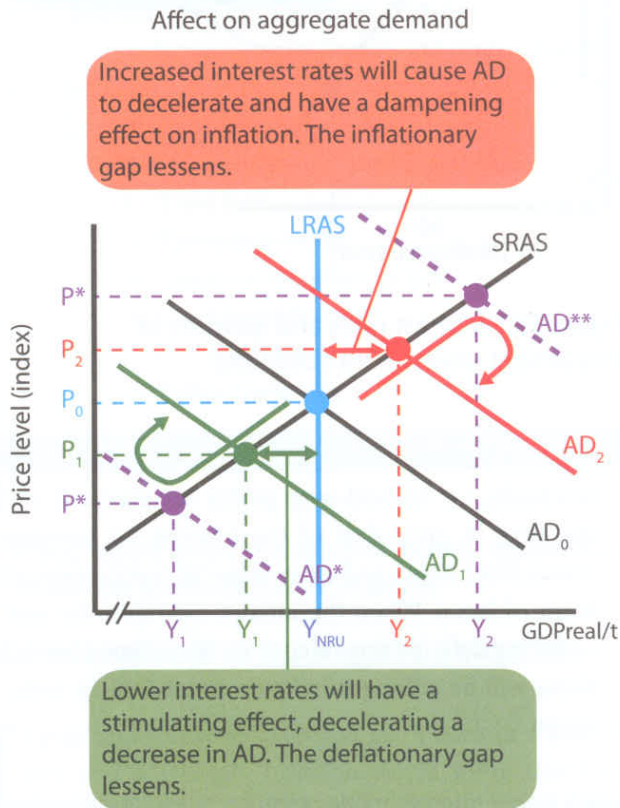
Affect on aggregate demand

> Increased interest rates will cause AD to decelerate and have a dampening effect on inflation. The inflationary gap lessens.

> Lower interest rates will have a stimulating effect, decelerating a decrease in AD. The deflationary gap lessens.

Figure 59.2 Demand management using interest rates/supply of money

Now, say that the economy shows signs of *overheating*, i.e. that the rate of inflation exceeds a predetermined ceiling rate of inflation set by the central bank. The central bank, implementing **tight monetary policy**, raises the discount rate to 3.5% whereupon commercial banks immediately respond by raising their lending rates to households and firms – since the commercial banks will now pay a higher 'price' for borrowing from the central bank. The deposit rate will also increase, since banks will be competing for households' and firms' deposits. Assume that the commercial banks increase the lending rate to 7.5% and the deposit rate to 2.5% - this will increase the opportunity cost of borrowing for consumption/investment and therefore increased savings and decreased consumption/investment. Taken together, lower levels of investment and

consumption will decrease aggregate demand, or, to be more correct, *decrease the rate of increase* in aggregate demand, as illustrated in the 'hook' between AD^* and AD_2 in figure 59.2.[4]

Note how the effect of an increase or decrease in interest rates is illustrated in Figure 59.2 as an 'increase/decrease in the rate of decrease' in aggregate demand rather than using a single aggregate demand curve that shifts to the right or left. This is not by accident. The reasoning is that monetary policy is often very 'forward looking' and seeks to countermand inflation that has not set in yet. Central banks attempt to steer the economy – or rather, inflation – by looking forward in time and estimating what inflation *will* look like – for example by using the PPI to gauge coming inflationary pressure (see Chapter 52). Interest rates are in effect used to countermand excessive changes in aggregate demand before they happen. This is because there are significant *time lags* in operation; it takes between two and six quarters before interest rates actually affect inflation and up to two years before the full effects on aggregate demand hit. This is why so much effort is put into predicting future output fluctuations in business cycles.

Another method of loosening monetary policy is by increasing the supply of money as explained in Chapter 58. This is illustrated in Figure 59.3 I-III following the same line of reasoning as above.

- *Increased money supply* (Sm_0 to Sm_1) puts downward pressure on interest rates (r_0 to r_1) and stimulates consumption and investment expenditure in the economy – note the use of the investment schedule – aggregate demand is stimulated from AD^* to AD_1
- A *decrease in the supply* of money (Sm_0 to Sm_2) causes interest rates to rise (r_0 to r_2) and causes aggregate demand to contract from AD^* to AD_2.

---

4 An additional effect is that when interest rises it may trigger an increase in demand for the Home Currency – which thus appreciates the Home Currency causing dearer exports and cheaper imports. Falling export revenue and rising import expenditure will act in the same direction as in falling consumption and investment; *decreasing aggregate demand.*

MACROECONOMICS

I: Supply and demand for money

II: The investment schedule

III: Effect on aggregate demand
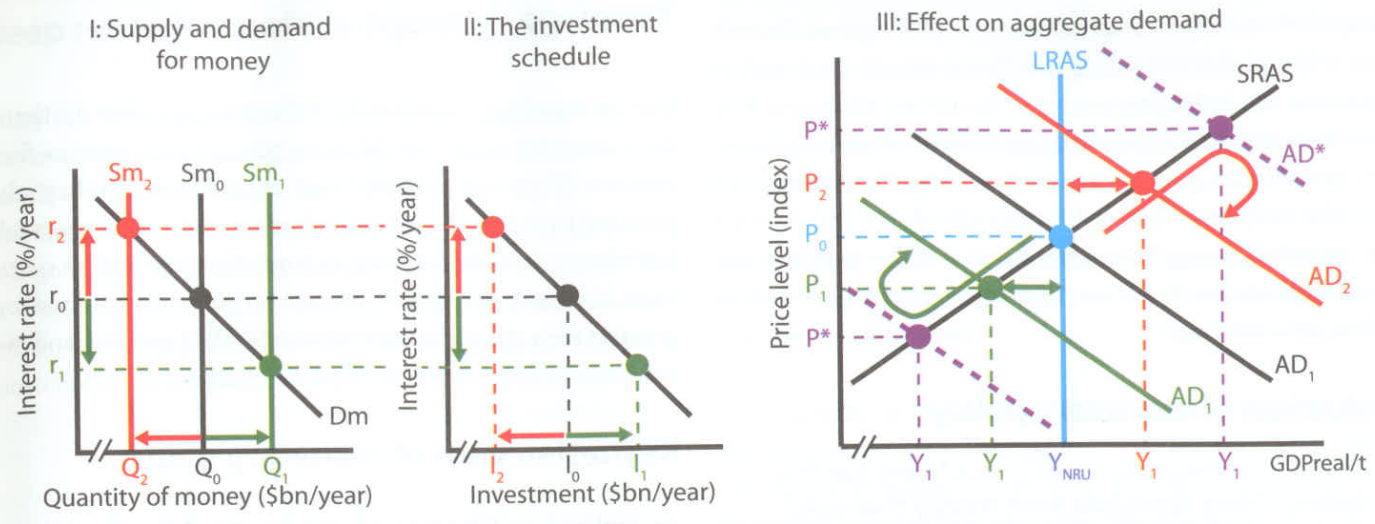
Increased supply of money.....→.....lowers the rate of interest.....→.....which stimulates C and I in AD.

Decreased supply of money.....→.....raises the rate of interest.....→.....which decreases C and I in AD.

Figure 59.3 Supply and demand for money linked to aggregate demand

Inflation targeting

> 'Goodhart was an optimist – just like Murphy!'
> Author's comment on Goodhart's law[5]

During the latter part of the 1900s, central banks gained increasing independence from governments in order to increase the predictability of central bank policies and also – if one is to be a tad cynical – to limit the meddling of politicians using loose monetary policy to fan the flames of their re-election. The idea is that when a central bank has a publicly displayed inflation target of, say, 2% then stakeholders such as firms and consumers can readily anticipate central bank policy moves since the measure of inflation (CPI in most cases) is publicly available on a day-to-day basis. If the central bank 'keeps its word' over a period of time, then households and firms can plan ahead in *anticipation* of higher or lower interest rates and this should even-out cyclical swings in the economy. This in turn should allow a steadier rate of growth and even induce greater investment as inflation rates become more predictable.[6]
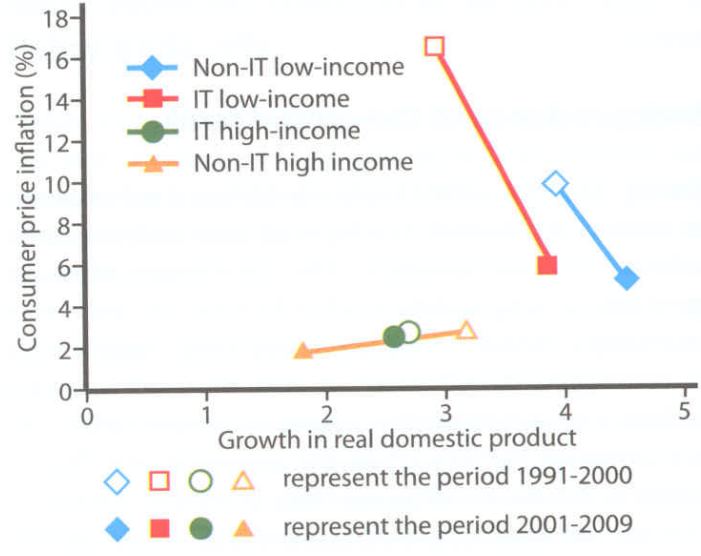
Figure 59.4 Inflation and growth in inflation-targeting countries

Source: *Inflation Targeting turns 20*, Scott Roger, IMF, at http://www.imf.org/external/pubs/ft/fandd/2010/03/pdf/roger.pdf

Some 26 countries in the world apply inflation targets with set inflation target rates of around 2 to 3% inflation. For example, one of the first countries to adopt an inflation target was Sweden, in 1993[7], where Svenska Riksbanken (the central bank) has had a target of 2% inflation per year with a +/- tolerance of a percentage point – e.g. inflation is allowed to fluctuate

---

5 Goodhart's Law states that any time there is an observed regularity – e.g. pattern – in statistical data and one tries to use the correlation for a given purpose, the observed pattern breaks down. Murphy's Law states that if something can go wrong, it will. I think this quote is from the end of my fifth marriage.

6 See a very accessible study by NBER at http://www.nber.org/papers/w16654.pdf

7 My country had just gone through a severe currency and banking crisis – just like the UK, which had adopted an inflation target the year before Sweden. Look up 'George Soros 1992' for an account of what happened.

between 1% and 3% without central bank intervention. There is some evidence that the setting of inflation targets does work to lower long run inflationary trends. Figure 59.4 illustrates how inflation-targeting (IT) countries generally had lower inflation rates and higher growth trends than non-IT countries. (It should be mentioned that the US central bank, the Fed, does not have an inflation target but instead a goal of low inflation and low unemployment. However, it seems to be leaning *towards* an explicit inflation target.)

Evaluation of monetary policy

> "If you have five dollars and Chuck Norris has five dollars, Chuck Norris has more money than you."
> Anonymous

Monetary policy comes with a rather mixed bag of successes, limitations and trade-offs. While one can generally claim that most countries have leaned more heavily on monetary than fiscal policies in later years, there is also no denying that there is a good reason why not all countries have adopted inflation targets.

Independence of the central bank

During the 1990s, central banks saw increased independence in terms of the mandate to achieve set goals such as growth, inflation and unemployment. The distortionary effects of governments using monetary policy for their own gains were increasingly viewed by academics and policy-makers alike as outweighing the advantages of using fiscal and monetary policies in a concerted effort to implement set policy goals. One key advantage put forward by the monetarist/new-classical school is that the crowding out effect is avoided. Another is that one can avoid a 'political business cycle' where different governments issue widely different policy objectives over time. Finally, it seems empirically reasonably clear that increased central bank independence results in a lower long-run level of inflation.[8] It should be pointed out that, while central bank independence has always been advocated by the monetarist/new-classical school, it has become a rather mainstream view.[9]

---

8 See for example a study from the University of California at http://people.ucsc.edu/~walshc/MyPapers/cbi_newpalgrave.pdf

9 See for example the view of the European Central Bank (ECB) at http://www.ecb.int/press/key/date/2007/html/sp070419.en.html

Time lags

Not all lags and concomitant difficulties in correctly timing contractionary and expansionary policies arise due to fiscal policies. There are notable **lags in monetary policy** also, pointing to at least 2 quarters before interest rate changes take full effect - and often as long as 6 quarters. The British central bank, the Bank of England, estimates that it takes at least four quarters for a change in interest rates to affect inflation and over two years before maximum impact is reached.

Keynesian view of monetary policy

As outlined in Chapters 45 and 46, the difference between the Keynesian and new-classical aggregate supply curves has implications for economic policy. A key element in the Keynesian/new-classical debate is whether monetary policy is more effective than fiscal policy. The Keynesian view remains sceptical of the effectiveness of monetary policy in adjusting aggregate demand by pointing out that an increase in the money supply does not necessarily mean banks will lend out the excess in reserves. Another point is that, while both consumption and investment are indeed linked to interest rate changes, they are interest-inelastic and thus there is limited impact on these two key components of aggregate demand. It is for these reasons that Keynesian economics focuses more on fiscal policy.

However ...

Monetarists generally retort that even given the existence of a liquidity trap – by no means certain, they say – monetary policy is not without teeth even at zero or near-zero interest rates. Increasing money supply can fuel expectations of households and firms, thereby fulfilling the criteria for expectations-based increases in aggregate demand.

Limits of monetary policy

Monetary policy has shown severe limitations historically in dealing with deep recessions and supply-side shocks. During the 1973-'75 period (the 'First Oil Shock') there were virtually no non-oil-producing countries spared the stagflation resulting from quadrupled oil prices. The aggregate supply shock in many cases led to a cost-push spiral (see Chapter 53) and unforeseen levels of unemployment. This seriously limited monetary policies since any attempts at contractionary monetary policies would have (at least initially) led to increases in unemployment that society would have found unacceptable.

Deep recession and the liquidity trap

A key Keynesian argument is the so-called **liquidity trap**. This is a situation where any increase in the supply of money has no effect on interest rates – the short term interest rate is zero. This negates any expansionary monetary policy as illustrated in Figure 59.5 where an increase in the supply of money (Sm_0 to Sm_1) has no effect on interest rates. Since interest rates remain unchanged there is no decrease in savings or increase in consumption/investment.
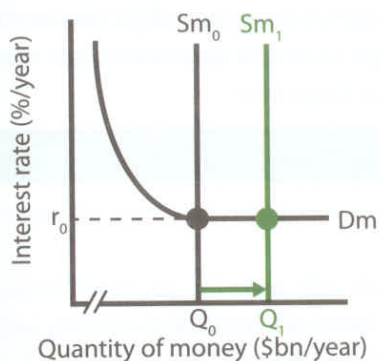
Increased supply of money.....→.....no effect on the rate of interest.....→.....no change in C or I in AD.

Figure 59.5 Liquidity trap

During deep recessions, central banks will commonly have lowered interest rates to fight possible deflation and fuel consumption/investment. The problem is that, when money supply has increased to the point where interest rates hit zero, there is nowhere left to go. My grandfather, bless his Missouri rough-neck soul, would have listened intently (beer and cigarette in hand) and said "Sounds like the injuns are still coming and you ran out of ammo, son!"[10]
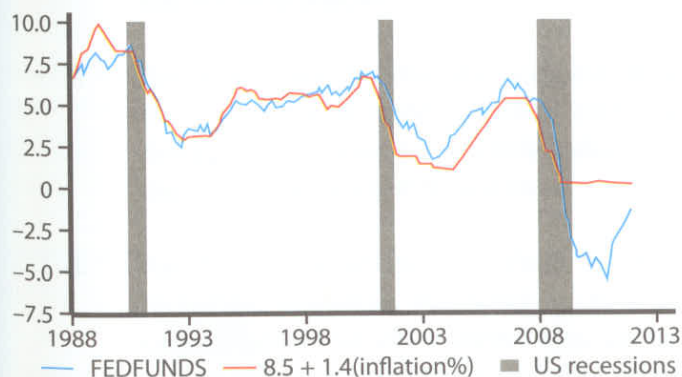
Figure 59.6 The 'Mankiw Rule of interest rates' and de facto Fed rates, 1988 to 2011

10 'Injuns' means Indians and 'ammo' is ammunition. My American-Irish family comes from the same patch of Missouri woods as the James and Younger brothers – bandits, brigands, killers and cut-throats to the last man.

An interesting 'test' or 'proof' of the limits of monetary policy during deep recessions was provided by professor Greg Mankiw at Harvard University at the start of 2012.[11] He had earlier written a simple formula suggesting that the Federal Reserve (the US central bank) set interest rates in accordance with:

Interest rate = 8.5 + 1.4(inflation – unemployment). This means that during high inflation and low unemployment, interest rates are raised.

For example, at the start of 2011, inflation was 2.1% and unemployment 8.9%. This gives 8.5 + 1.4(2.1-8.9); the interest rate should be minus 1.02 percent! Clearly not possible. This is illustrated in figure 59.6 where the 'Mankiw Rule' follows the de facto Fed rate very closely. That is, until the advent in 2008 of what would become a severe recession. The discrepancy shows that the Fed rate basically hit zero while the 'Mankiw Rule' suggested far lower rates – e.g. *negative* rates of interest. The US hit a liquidity trap.

Policy trade-offs

Finally, more trade-offs. In Chapters 48 and 57 five key trade-offs were identified, such as growth and inflation, inflation and unemployment and domestic monetary policy and stability in the exchange rate. These are all still relevant in the field of monetary policy:

- *Tight* monetary policy ($\uparrow\Delta r$ or $\downarrow\Delta Sm$) aimed at...

  - ... decreasing inflation. This will contract economic growth and fuel unemployment. It can also increase demand for the domestic currency and drive up the exchange rate.
  - ... appreciating the Home Currency. The trade-off is of course lower growth and higher unemployment.

- *Loose* monetary policy ($\downarrow\Delta r$ or $\uparrow\Delta Sm$) aimed at...

  - ... increasing growth and/or decreasing unemployment. This can lead to a decrease in demand for the Home Currency and a depreciation of the exchange rate.
  - ... depreciating the exchange rate in order to increase exports and/or decrease imports. The effect will be inflationary as both consumption and investment fuel aggregate demand.

11 See Professor Mankiw's page at http://gregmankiw.blogspot.com/2012/01/liquidity-trap-may-soon-be-over.html

POP QUIZ 59.1

Demand-Side Policies

1. Explain how inflationary expectations in households may influence aggregate demand.

2. How are the variables M, Y, CPI and transfer payments commonly affected during a recession?

3. Explain how fiscal policy may be used in an economy showing increasing inflation.

4. Draw four separate AS-AD diagrams (and perhaps supporting diagrams!) showing the effects of:

   a. lower income taxes,

   b. increased interest rates

   c. increased supply of money

   d. higher exchange rate.

5. Explain why an increase in aggregate demand caused by increased government spending may not increase real output in the long run. Use the AS-AD model in your answer.

6. "The government raised interest rates today in response to a fiscal stimulus package presented by the Central Bank." What's wrong with this picture? Look carefully.

7. Why may aggregate demand fall as a result of falling property prices? Why may a decrease in aggregate demand result in falling property prices?

8. Canada and Mexico are trade partners. If Canada's inflation rate increases at a faster rate than Mexico's, *ceteris paribus*, how may one expect aggregate demand to be affected in each country?

9. A tricky one; in an economy the interest rate falls from 9% to 7% and at the same time the rate of inflation falls from 8% to 5%. Explain why borrowers may in fact not be better off.

10. Explain how the Central Bank could reduce the rate of inflation in the economy.

11. Using the AS-AD model, evaluate the effects of monetary policies aimed at reducing inflation.

12. Using an appropriate diagram, explain why it may be very difficult for a central bank to stimulate aggregate demand in times of severe recession.

13. Thinking ahead to Section 3 (trade): Assume that a decrease in the exchange rate (depreciation) has shown to have a positive influence on aggregate demand in an economy due to a rise in export revenue and/or a decrease in import spending. Might the inverse be true also, e.g. might increased national income instead affect a country's exchange rate?

Summary & revision

1. The **real interest rate** is defined as the nominal interest rate minus inflation.

2. **Interest rates** and **investment expenditure** (fixed capital) are negatively correlated.

3. **Loose monetary policy** is either decreasing interest rates or increasing the supply of money. Both policies have an expansionary effect on aggregate demand, since lower interest stimulates consumption and investment.

4. **Tight monetary policy** is raising interest or decreasing the supply of money. This has a contractionary effect on aggregate demand as consumption and investment fall.

5. **Inflation targeting** is when the central bank clearly states a given level of aimed-for inflation over a period of time. Since inflation figures are openly available it allows stakeholders (firms and households) to plan ahead and anticipate central bank actions.

6. **Critique** of monetary policy:

   a. The degree of **central bank independence** influences monetary policies. Generally, the freer a central bank is to set policy, the greater the economic stability.

   b. **Time lags** in monetary policy can have the effect of exacerbating cyclical variations in the economy, just like fiscal policies.

   c. Keynesians are traditionally sceptical about the general use of monetary policy, citing that demand for investment is relatively inelastic and thus insensitive to lower interest rates.

   d. Another Keynesian point is that, in severe recessions, economies can wind up in a so-called liquidity trap where interest rates are zero and increased money supply therefore cannot stimulate increased consumption and investment.

e. Monetary policy comes with the same trade-offs as in fiscal policy, namely

   i. $\downarrow\Delta r \rightarrow \uparrow\Delta Y \Leftrightarrow$ price stability and depreciation of the currency

   ii. $\uparrow\Delta r \rightarrow \downarrow\Delta$inflation $\Leftrightarrow$ growth and higher price of exports

MACROECONOMICS

413

2.6

60. The Role of Supply-Side Policies

Key concepts:

- Definition of supply-side policies
- Categories of supply-side policies
- Market based and interventionist policies

"I've talked to you on a number of occasions about the economic problems our nation faces, and I am prepared to tell you it's in a hell of a mess…we're not connected to the press room yet, are we?" (Ronald Reagan)

> **Definition: 'Supply-side policies'**
> Policies aimed at increasing the long run potential output in an economy, such as deregulating labour markets, lowering corporate taxes and subsidising infrastructure, are known as supply-side policies. The aim is to improve overall productive capacity in an economy.

Definition of supply-side policies

Check back to Chapter 44 where I state that, since there is no correlation between LRAS and price, there are serious implications for economic policies. Well, here they come. The upward-sloping short-run aggregate supply curve (SRAS) is positively correlated to the price level – as output increases firms hit diminishing returns and bottlenecks in supply. SRAS *shifts* due to changes in factor availability and factor prices – primarily changes in the price of raw materials, wage rates, labour market legislation and taxes linked to labour/capital.

The long-run aggregate supply (LRAS) curve is perfectly inelastic, e.g. there is no correlation between the price level and output. LRAS *shifts* when the **quantity** and/or **quality of factors** of production change. Policies implemented by governments to structurally change the productive potential of an economy (similar to that of shifting the PPF to the right) are known as 'supply-side policies'.

Categories of supply-side policies

Supply-side economists favour polices aimed at creating the basic economic conditions for long-run increases in output, where **long run aggregate supply increases** over time, allowing aggregate demand to increase without creating *only* inflationary pressure. (See Figure 49.2 in Chapter 49). The increase in LRAS allows for increased demand without creating harmful inflation rates which would decrease short-run aggregate supply and *potential* (full employment level) output increases over time. Supply-side policies in fact aim for the same thing as demand side; increased growth, lower unemployment, price stability and equilibrium in the foreign sector. It is therefore not the *aim* which differs but the *method* of achieving these aims. Policies aimed at increasing LRAS aim to:

1. Provide **incentives** for labourers to work more and unemployed to search for jobs

2. Increase the **mobility of labour** – both between jobs and between regions

3. Increase labour and capital **productivity**

4. Increase firms' ability and willingness to invest, e.g. **increase the physical capital** stock

5. Foster **innovation** both in production and products

6. Increase **competition** in the economy via privatisation and ease of start-ups

7. Increase **foreign direct investment** (FDI) – both abroad and at home

All the above aim to increase growth over the long run without experiencing the trade-offs common to demand-side policies.

Market based and interventionist policies

Supply-side economics arose out of the same mould as the monetarist/neo-classical school during the 1980s and focuses on policies which enhance the long-run output potential in the economy by way of creating well-functioning factor markets and incentives for both producers and labourers. Supply-side policies are in essence **microeconomic** policies, since the policies target specific markets such as the labour or capital market.

Market based policies: Supply-side policies are frequently highly *market orientated*, aiming to 'liberate' markets which are hindered from clearing due to various forms of market 'impurities' such as labour legislation, strong unions and harmful government tax laws. The US and UK implemented many market based policies in the 1980s.

Interventionist policies: There are, however, various forms of interventionism such as increased government investment in infrastructure and education, government re-skilling and training schemes and regional/sectoral support in depressed areas/sectors. Scandinavian countries were noticeably active in this area during the 1990s.

Summary & revision

1. **Supply-side policies** are government policies aimed at increasing the long-run potential of an economy, e.g. shift LRAS.

2. **Categories** of supply-side policies include labour market reforms aimed at clearing the labour market better, tax policies designed to increase firms' willingness/ ability to produce, competition policies enabling greater competition, and policies designed to increase innovation.

3. **Market based supply-side policies** focus on getting markets to function better by removing market impediments such as minimum wages, too-high taxes and government regulations which stifle competition.

4. **Interventionist** policies create the preconditions for long-run growth via government provision or encouragement of education, increased infrastructure and improved information and assistance in job-brokering.

61. Interventionist Supply-side Policies

Key concepts:

- Investment in human capital
- Policies encouraging investment (physical capital)
- Infrastructure
- Industrial policies

Supply-side policies are not the sole domain of neo-classically orientated free market policies, yet care should be made to distinguish *market*-based policies from *interventionist* ones. The 1980s and '90s saw increasing use of government empowerment to influence the supply side of the economy. Generally speaking, governments intervened on factor markets by trying to enhance the attractiveness and availability of labour, i.e. the supply of labour, which also had the secondary effect of increased demand for labour over time.

Investment in human capital

Basically, any government policy causing an increase in labour market participation increases the supply of labour. Figure 61.1 shows how various forms of *interventionist supply-side policies* – notably education, (re-)training and greater labour mobility – increase the aggregate supply of labour. This in turn increases the potential output in the economy, LRAS.
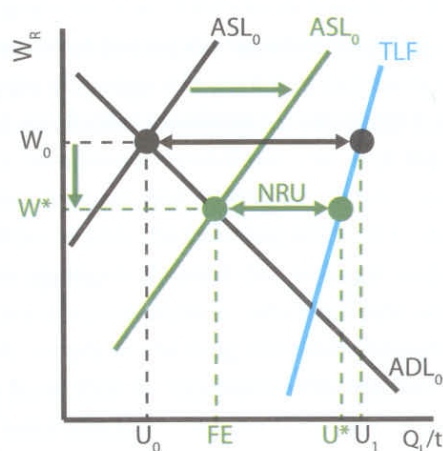
Government intervention on the supply-side increases the aggregate supply of labour from ASL_0 to ASL_1. As real wages fall, there is an increase in quantity of labour demanded and unemployment falls from $U_0 \Leftrightarrow U_1$ to $FE \Leftrightarrow U^*$ - the natural rate of unemployment.

Figure 61.1 Intervention on the supply-side for labour

By increasing the skills base of labour, decreasing the search costs of both employers and job searchers and creating incentives for increased labour mobility, there will be an increase in the aggregate supply of labour, shown in Figure 61.1 as the shift in the aggregate supply of labour from ASL_0 to ASL_1.[1] Assuming that wages adjust relatively quickly – e.g. no downward stickiness of wages – the real wage rate falls to W^* and unemployment decreases from $U_0 \Leftrightarrow U_1$ to $FE \Leftrightarrow U^*$ - the natural rate of unemployment.

Investment in human capital as a result of government intervention (see Chapter 49) takes on some of the following range of interventionism:

- Government *grants/subsidies* to employers hiring youths, older workers and long term unemployed

- Entrepreneurial incentives such as *soft loans* and subsidised rent for start-up companies and R&D loans

- Government *tax incentives* to firms which invest in education/training amongst employees risking redundancy

- Government/communal *re-training schemes*

- *Improved information* for job-seekers by way of government employment agencies, on-line job seeking and help/advice in job seeking

- *Regional support* and outsourcing of government agencies to depressed areas

- Specific *youth training programmes* and subsidies for firms hiring under-18s (which will give these young people much needed experience to gain 'real' jobs)

1 It is highly plausible (= believable) that these same policies, together with hiring incentives for firms, may ultimately also result in an increase in *demand* for labour.

It bears commenting on the rather weak conclusions as to whether government re-training and education programmes in fact have much of an effect! Recent findings indicate that, while focused, small-scale government training schemes may have some effect, there is no general rule that increased government spending on training schemes has any significant effect on unemployment and long-run output.[2] Having said that,

Figure 61.2 corroborates this to a certain extent, where several of the Scandinavian countries (Norway, Sweden and Denmark) using large-scale government supply-side policies on the labour market generally have a better unemployment record.[3]

---

2 See for example *The Role of Training and Skills Development in Active Labour Market Policies*, Institute for Employment Studies, University of Sussex, 2009

3 Though the regression coefficient is less than 0.4.
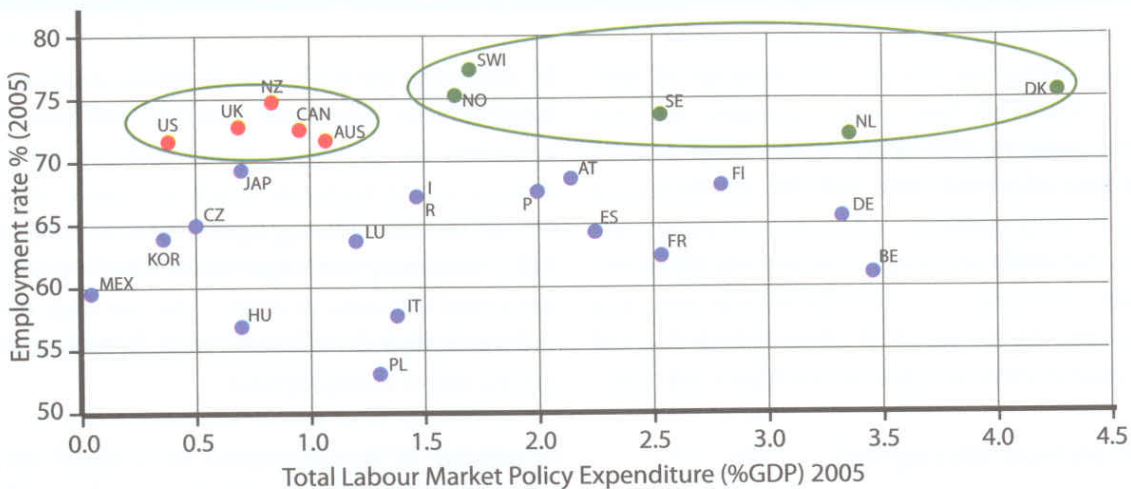
(Source: "The Role of Training and Skills Development in Active Labour Market Policies", Institute for Employment Studies, University of Sussex, 2009, page 6.)

Figure 61.2 Correlation between labour market spending and employment (2005)

Policies encouraging investment (physical capital)

Interventionist policies to encourage investment and start-up firms commonly include such measures as:

- Subsidies for start-up firms

- Tax breaks for firms investing in physical capital

- Tax holidays (as in Ireland) for foreign firms investing in the economy (FDI inflows)

- Soft loans (loans with lower-than-market interest and/ or longer repayment times) for firms needing to invest.

All too often students see the similarities between simple demand (e.g. demand for a single product) and aggregate demand (e.g. the sum of all product markets) and draw the conclusion that – as in the simple supply and demand model – aggregate demand does not influence aggregate supply. This is in fact erroneous, since it is quite possible that the investment

component of aggregate demand indeed will serve to cause a change in both short-run and long-run aggregate supply over time. Figure 61.3 illustrates an economy initially in long-run equilibrium, point A.
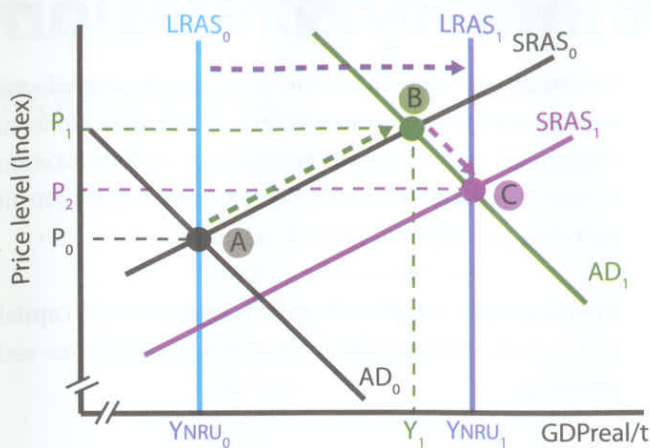
- **Investment expenditure increases AD (AD$_0$ to AD$_1$)**

- **Physical capital used in production shifts SRAS (SRAS$_0$ to SRAS$_1$)**

- **New LR equilibrium established at Y$_{NRU_1}$ (LRAS$_0$ to LRAS$_1$)**

Figure 61.3 LRAS effects due to increased physical capital

A to B: Investment increases and AD shifts from **AD$_0$** to **AD$_1$**, taking the economy from point **A** to point **B**.

B to C: In due course, the effects of increased capital will increase productivity, causing aggregate supply to increase from **SRAS$_0$** to **SRAS$_1$**; point **C**.

Increase in LRAS: The increase in productivity and output capability causes LRAS to increase from **LRAS$_0$** to **LRAS$_1$**, creating new long run equilibrium at Y$_{NRU_1}$.

This is a case where a change in a component of aggregate demand, *investment expenditure*, can directly influence both aggregate demand and aggregate supply over time. Other possibilities include the lowering of income taxes which at first increases aggregate demand due to larger disposable incomes and increased consumption, which in turn serves to increase aggregate supply in the long run when households' incentives to work more kick in. Increased infrastructure investment also has the same effect.

Infrastructure

When economists use the term infrastructure, they are referring to the 'internal skeleton' which holds an economy upright and moving. Infrastructure provides the basis for an economy to function, and consists of the system of road networks, telecommunications networks, sewage facilities, water supply,

electricity and power sources, financial/education/health system, public transportation networks and ports/harbours or other international points of trade access. An increase in infrastructure will (as in the examples referring to figure 61.3) initially increase aggregate demand.

Infrastructure is vital to a well-functioning economy and the effects of investment here are hard to over-estimate. It allows all the factors of production in a country to be productively utilised: labourers can get to their jobs; produce can be transported to urban areas from rural areas; power, water and information so vital to firms can be transported; and enhanced market accessibility increases the level of competition in an economy.

Allow me a moment of patriotic breast-beating for my wonderful home country of Sweden. The Swedish government has consistently and generously built up a telecommunication structure of broadband which basically grants the entire population access to high-speed internet. This has made Sweden the most competitive digital economy in the world according to both The Economist's *Digital economy ranking 2010* and the World Economic Forum's *Global Information Technology Report 2009-10*. This seriously switched-on and hooked-up little country of 9.3 million people has quite a few global success stories to its credit such as Spotify and Skype. Most of the Swedes would agree that large-scale internet access and focused IT classes in school together with government investment in infrastructure is at the heart of the success.

Industrial policies

Finally, it is common for governments to take an active role in helping industries via government legislation and allocation of funds. The following industrial policies can increase long-run potential:

- R&D grants to firms, together with government support for R&D units linked to state universities

- Regional support for fledgling industries

- Relocation subsidies for workers to seek and take jobs in other regions

- Active promotion of Home Country industries and products via trade legations at embassies abroad.

MACROECONOMICS

POP QUIZ 61.1

Supply-Side Issues

1. Show in a diagram how supply-side policies may lead to deflation. What might cause this and is it 'bad'?

2. Why do neo-classical economists claim that demand-side policies do not increase real long-run output? Illustrate your answer with a suitable diagram.

3. Explain how supply-side policies can be applied to the labour market.

4. What would the supply-side effects be of an increase in a) income taxes; b) tax breaks for small businesses c) fewer subsidies to 'sunset industries' (industries experiencing permanent decline in demand for their goods) in regions with high unemployment rates?

Summary & revision

1. Governments can intervene in **human capital** via subsidies/grants/tax breaks to firms which train workers; implement government re-training and education schemes; centralised job centres and employment agencies; and regional support schemes.

2. Interventionist supply-side policies for **physical capital** can consist of tax breaks for investment, soft loans and subsidies.

3. Governments can enhance long-run potential output by increasing the quality and quantity of **infrastructure** such as roads, bridges, ports, telecommunications and water supply.

4. **Industrial policies** can actively promote R&D for private firms and state universities and aid the relocation of firms and workers from sunset to sunrise industries.

62. Market-based Supply-side Policies

Key concepts:

- Encouraging competition
- Labour market supply-side policies
- Incentives
  - Labour market incentives
  - Investment incentives
- Evaluation of supply-side policies
  - Time lags
  - Limitations of tax cuts
  - Budget deficits
  - Social costs
  - Imperfectly functioning markets

The main theme in supply-side policies is that market forces are far better at creating output in the long-run than government intervention in the form of demand management. *Sustainable* – long-run – increases in real GDP are only possible, according to this neo-classically orientated view, by increasing long-run aggregate supply. Allowing firms to make a profit under competitive forces and creating conditions for factor markets to clear will shift the long-run aggregate supply curve and therefore decrease the natural rate of unemployment. Market-based policies aimed at the supply-side in effect seek to 'restore' the basic function of the price mechanism by removing government involvement which distorts markets and creates disequilibrium.

Encouraging competition

Supply-side economists point to the importance of *competition* in an economy as an overriding element in increasing long-run aggregate supply. To this end, the following competition policies are put forward:

- *Privatisation* of government-run businesses and deregulation of markets are staple supply-side measures in increasing competitive forces in an economy. This has often been done in conjunction with government spending cuts.

- *Encourage entrepreneurship* by granting 'tax holidays' and creating beneficial funding schemes for start-ups, e.g. new firms.

- *Trade liberalisation* – reducing tariffs (= import tariffs) and other barriers to free trade – and free capital flows (easier foreign investment) are policies often put forward by supply-siders.

Labour market supply-side policies

Supply-side policies in labour markets centre on increasing labour mobility – both in terms of geographical location, industry and 'time between jobs' – by increasing the incentives of firms to hire and workers to accept jobs. The basic ideal is to increase the supply and overall quality of labour while creating mechanisms for well-functioning labour markets. Market based policies can consist of:

- Change labour laws and make it easier for firms to *hire and fire*, reducing search costs and risks for firms while decreasing between-job time for workers.

- *Reducing/abolishing regional support schemes* in order to highlight differences in regional unemployment levels and thus encourage people to move to new jobs in other regions.

- *Cutting back on social welfare*/unemployment benefits in order to encourage people to accept jobs by increasing the opportunity costs of unemployment.

- *Reducing union power* – say, by making sympathy strikes illegal and making collective bargaining harder – in order to reduce wage stickiness.

MACROECONOMICS

421

- *Abolish minimum wages* to allow market forces to set wages.

By lowering *union power*, lowering *taxes on labour*, and decreasing *minimum wages* it is possible to reduce the influence of labour market imperfections which diminish the ability of the market to clear (see Chapter 51, figure 51.6). If successful, the real wage rate will conform to labour market forces and lower the wage rate from W_0 to W^* (figure 62.1) decreasing total unemployment from $U_0 \Leftrightarrow U_1$ to the natural rate of unemployment of $FE \Leftrightarrow U^*$.
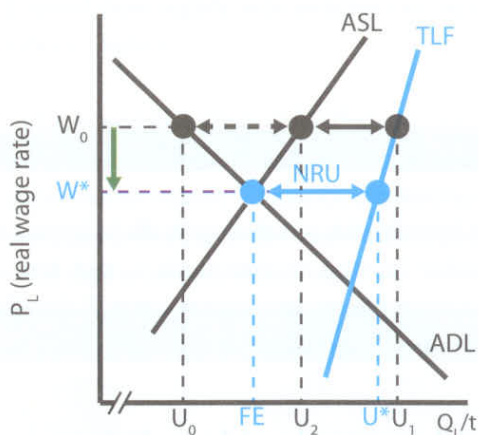
The removal of labour market impurities (minimum wage and union power in wage negotiations) puts downward pressure on real wages (W_0 to W^*) and the labour market clears at the natural rate of unemployment, $FE \Leftrightarrow U^*$.

Figure 62.1 Removing labour market imperfections – a fall in real wages

OUTSIDE THE BOX

Decreasing the natural rate of unemployment

Other supply-side policies for labour aim to reduce labour market rigidities and increase the supply of labour by increasing the overall propensity of workers to accept jobs. Assume an economy in labour market equilibrium with a natural rate of unemployment of $FE_0 \Leftrightarrow U^*_0$, Figure 62.2. By implementing various *supply-side measures*, the aggregate supply for labour

increases, i.e. more people in the labour force are willing to supply their labour to the market at all real wage levels. This is shown in a shift in aggregate supply of labour from ASL_0 to ASL_1, which lowers the natural rate of unemployment from $FE_0 \Leftrightarrow U^*_0$ to $FE_1 \Leftrightarrow U^*_1$.
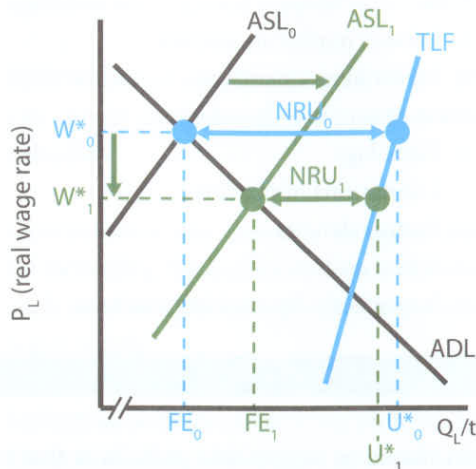
Labour market reforms can increase the willingness of labourers to accept jobs at all real wage levels, e.g. an increase in the aggregate supply of labour (ASL_0 to ASL_1). The natural rate of unemployment falls from $FE_0 \Leftrightarrow U^*_1$ to $FE_1 \Leftrightarrow U^*_1$.

Figure 62.2

Such policies, commonly referred to as **structural reforms**, often include:

- *lowering social/unemployment benefits* in order to persuade labourers to accept jobs (since lower benefits increase the opportunity costs of remaining unemployed)

- *lowering income taxes* to induce increased labour hours and create an incentive for those in the labour force to accept jobs

- *increasing overall labour flexibility* by reforming labour market legislation, for example by making it easier to hire/fire labourers and lowering mandatory severance pay

- using *retraining and education* schemes to enable labourers to decrease time spent between jobs and to improve reallocation of labour from declining industries to growth industries.

The overall aim of market solutions to unemployment is to alleviate supply and demand mismatches, i.e. to improve labour allocation. Proponents of market solutions point to the US and Great Britain as examples of how long-run unemployment rates fell markedly during and after the Reagan and Thatcher reforms of the 1980s.

A Little Depth

Tax incentives

"Well, of course, when you cut taxes, government revenues go up. Why couldn't I see that before?"

Time columnist Michael Kinsley's first words – according to his editor – upon waking up from brain surgery.

Decreasing various taxes are forms of competition policies since the objective is to increase workers' propensity to work and firms' propensity to produce. Taxes tend to skew markets (see Chapter 13) and create misallocation, so reducing/removing taxes should result in improved resource allocation and increased output. Again, it should be noted that many taxes initially affect aggregate demand before affecting short-run aggregate supply and long-run aggregate supply.

- Tax breaks/deductions to firms for (re-)*investment* will stimulate investment expenditure.

- Lower taxes on *dividends* (a share of company profits paid to shareholders) and other forms of capital gains taxes can increase investment funding for firms since more people will be willing to buy shares and bonds.

- Lower *profit* (corporate) taxes encourage further investment by firms.

- Cutting *labour* taxes (called payroll taxes in the US) lowers costs for firms.

- Decreased *marginal tax* rates on income act as an incentive for labourers to work more; and decrease labour taxes in order to decrease labour costs for firms.

Laffer Curve

A famous supply-side economist during the Reagan administration in America, Arthur Laffer, suggested that it is quite possible/probable that in some cases a decrease in the tax rate would lead to an increase in tax revenue[1]. Figure 62.3 shows the Laffer curve, which illustrates the marginal tax effect on income and thus tax revenues; if the tax rate is higher than t^*, lowering income tax will increase tax revenue, while a tax rate below t^* means that tax revenue will increase if tax is increased.
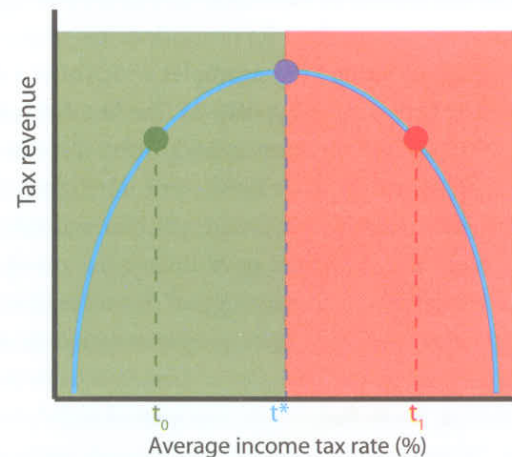

Figure 62.3 The Laffer curve

The argument supported by the Laffer curve is that high income taxes distort labour markets and lead to a socially sub-optimal level of labour hours and thus tax revenue. A cut in the (marginal) tax rate will mean an increase in real disposable income for labourers. Will people then increase their labour hours? Yes, according to supply-side economists, since this view asserts that the substitution effect is stronger than the income effect and therefore workers will have an incentive to work more. If a worker is earning €20,000 a year and paying 30% tax, then total tax revenue is €6,000. Say that the tax rate is lowered to 27% and that this induces the worker to increase his/

MACROECONOMICS

---

1 Laffer apparently drew the original curve on a cocktail napkin at a Washington restaurant on Dec. 4, 1974, to explain things to a journalist, Jude Wanniski of the *Wall Street Journal*, and future American Vice President *Dick Cheney*. See http://www.fortune.com/fortune/articles/0,15114,447776,00.html

her hours, increasing gross wages to €23,000. This would result in total tax revenue of €6,210.

Referring to the Laffer curve in Figure 62.3, if the income tax rate is above t* then total tax revenue will increase if the tax rate is lowered.

The Laffer curve suggests that there is an *optimal* tax rate in terms of optimising total tax revenue, t* in Figure 62.3. Naturally, tax revenue is zero at a tax rate of zero and would be zero at a tax rate of over 100% - since one would basically have to pay the government to work. Tax revenue will be collected in between these two extreme values and the general shape of the curve is hardly up for debate. The problem is in finding the point where taxes can be optimised to yield the highest revenue. I venture to say that while Laffer himself suggested that big governments and thus big taxes would in all likelihood be to the right of point **t*** and would benefit from lowering taxes, most economists would read purely empirical evidence to the effect that *high income* countries lie somewhere to the left of point **t***.

I cannot help mentioning a few examples which offer a *degree* of support for the Laffer curve hypothesis. The first is from Russia, where in 2001, President Putin introduced a flat rate income tax of 13% - down from over 30% - and during the first two years of the new (entirely non-progressive) tax rates income tax payments more than doubled government tax revenues. The tax enjoys seeming widespread support from Russians, to the extent that a survey showed more people in favour of a flat rate tax than a progressive tax.[2] On a more frivolous note, the Italian Prime Minister, Silvio Berlusconi, was quoted in February 2004 as saying: "If reasonable taxes are demanded, no one thinks about avoiding paying them. But if you ask 50% or more ... I consider myself morally justified to do everything I can to avoid paying them".[3] A more recent – still rather anecdotal – example comes from Bolivia where the effective tax rate on profit is 80% and tax revenues are far lower than in neighbouring Chile where profit taxes are considerably lower.[4]

---

2 Business Week, May 26 2003; *From each according to ... oh, never mind.*
3 *Time*, March 1 2004, page 17 (I must mention that the Italian budget deficit at the time of writing [January 2009] is circa 2.5% of GDP and that total national debt is 104% of GDP – the third highest in the world. See http://www.reuters.com/article/bondsNews/idUSLO27400220081125)
4 The Economist, *It's a jungle out there*, Oct 22, 2011

Evaluation of supply-side policies

It will come as no surprise that supply-side advocates are found primarily to the political right. The highest-profiled proponents of supply-side economics were Prime Minister Margaret Thatcher in Great Britain (1979 – 1990) and US President Ronald Reagan (1981 – 1989). Thatcher's government set out on a markedly supply-side program of privatisation and deregulation during the early 1980s when Britain was in severe recession. Trade union power was successively diminished, income taxes and marginal tax rates were drastically lowered and state-owned enterprises were privatised. The policies of Reagan (later dubbed 'Reaganomics') during the first half of the 1980s were similar; taxes were cut, government expenditure was slashed (creating large budget deficits and government debt) and government agencies downsized, and he implemented tight monetary policy to curb inflation. Reagan also increased military spending enormously. This, together with some of the largest tax cuts in American history, increased demand and helped bring the US out of the severe recession of 1981.

Criticism of supply-side labour policies

Critics of the Reagan/Thatcher economic regimens maintain that the social suffering of the supply-side reforms – such as marked increases in unemployment initially, increased income inequality, cutbacks in government programs and deteriorating social welfare nets – were excessive. While the final outcomes of both the Reagan and Thatcher eras remain highly debated, there is no doubt that in both cases severe inflation was brought under control and that productivity gains produced growth rates which were some of the highest in the industrialised world. The period also ushered in a more broad-based acceptance for 'rightist' policies based on supply-side measures – in fact, most western economies implemented various forms of supply-side measures during the 1990s.

Many of the weaknesses of supply-side policies put forward deal with the negative side effects on society. The list of supply-side policies given is ample indication of the possible effects on households comprised of low-income earners, people with only basic education, and older workers who would have difficulty switching jobs. Critics of supply-side labour market policies point to the following:

- **High social costs:** Lowering social/unemployment benefits and reducing minimum wages will have the severest effect on low wage earners and increase income inequalities in society. Supply-side policies also take far longer to have an effect on the economy than demand-side policies.

- **Limited impact of marginal tax rates:** Studies indicate that there is little effect on labour supply due to decreases in marginal income tax rates.

- **Evidence of demand-deficient unemployment:** Finally, observers are quick to criticise the neo-classical view that a large portion of disequilibrium unemployment is the result of 'mismatches' in the labour market which in fact implies that employers would increase their search for labour during recessions. This is belied by the fact that job postings and 'help wanted' ads fall markedly during economic downturns. This strengthens the case for at least some of the employment indeed being demand-deficient.

In spite of these 'grin-and-bear-it-policies', as Keynes called them, few critics of supply-side policies would dispute that long-term unemployment rates dropped substantially in the US and Britain during the 1980s and '90s. Instead, criticism is levied at a number of noticeable weaknesses of supply-side policies. Perhaps foremost is the implicit assertion that disequilibrium unemployment is primarily the result of labourers not being willing to accept jobs. Many economists would take issue with the strict categorisation of what is to be considered voluntary and/or frictional unemployment; a skilled piano builder can hardly be 'unwilling' or 'voluntarily unemployed' for failing to accept an available job waiting tables.

Incentive function of tax cuts is limited

While most economists accept that positive correlation exists between marginal income tax rates – the tax paid on the 'last' money earned – and the supply of labour, studies show a rather weak link. For example, the large tax cuts given in the US during the Reagan administration are estimated to have increased labour supply by between 0.4% and 0.9%[5]. Another argument put forward by supply-siders is that reductions in taxes and social benefits have a strong incentives effect and therefore increase productivity in the economy. This has been strongly contested by many Keynesian economists who point to the fact that labour productivity growth was virtually the same in the ten years before and after Reagan's supply-side measures.

Time lags

Critics of supply-side measures do acknowledge that cuts in taxes which affect output, such as profit and labour taxes, have had some measure of success in supply-side terms. However, they hasten to add that such policies take a long time to

affect the economy – 10 years or more – during which time a deep recession such as the 2008 global recession can cause considerable opportunity costs, such as…

Social costs and budget deficits

Demand management has often been used as a 'buffer' (= shock absorber) against alarming unemployment rates in various industrial sectors and geographical areas. Targeted increases in government spending, transfer payments and social programs have been used both in increasing aggregate demand and in 'social engineering', i.e. spreading wealth and curtailing income inequalities. Supply-side tax cuts have tended not only to increase **budget deficits** but also to starve the government sector of funds, which can lead to negative effects for poorer households when transfer payments and social programs diminish. Therefore governments run the risk of increasing **unemployment** in the short-run since the supply-side policies will initially have the effect of lowering aggregate demand, causing great economic and social loss during the time it takes for the supply-side policies to take effect.

The long-run effect is increased **inequality in income distribution**, something that one of the most vocal Keynesians, Paul Krugman[6], has pointed out in America, simply stating that "…Reaganomics made things worse, pushing millions of people… over the poverty line…"[7]. Critics also point out that diminishing the size of government and its spending is not only difficult and politically dangerous but can have negative effects on the long-run potential of the economy. Cost cutting in schools, health care and social programs proves to be economically foolish as it will adversely affect the long-run potential of the economy.

Imperfectly functioning markets

Keynesian economists generally view markets as imperfectly functioning mechanisms, where intervention by government is seen as well-motivated and necessary to correct some of the more glaring market failures. **Disequilibrium unemployment** – i.e. employment rates below full employment – is seen as the result of demand deficiency (see Chapter 51) and wage stickiness rather than the neo-classical view that market impediments simply do not allow labour prices to fall. In this manner, Keynesians argue that while the labour market *might* ultimately clear and return the economy to full employment, there would be considerable economic loss in the time being as factors lay idle and output was below potential. The correction

5 Abel & Bernake, page 578

6 Nobel Laureate in Economics 2008

7 1993 article by Krugman, published at; http://www.pkarchive.org/economy/ConservativeMirage.html

of demand-insufficiency in the economy lies in the hands of government interventionism and demand management.

Finally, a few words on the alluring – and erroneous – conclusion that economists who favour demand-side cuts are against all forms of supply-side policies. In fact, one of the areas of consensus in economic policy today is that supply-side policies indeed increase long-run aggregate supply. However, it is in the *choice* of supply-side policies that disagreement is still to be found. Neo-classical economists favour policies based on **free markets and incentives**, such as lower taxes and privatisation. Keynesian economists prefer supply-side policies of **interventionist type**, such as government funding of worker recruitment/education and re-training centres.

Preparing for exams

Short answer questions (10 marks each)

1. Why might there be unemployment at the equilibrium level of income?

2. Discuss whether inflation or deflation is the more serious problem for an economy.

3. Explain what the "natural rate of unemployment" is.

4. Explain how an unexpectedly large increase in the price of oil might affect a non-oil-exporting country.

5. What are the economic and social costs of high inflation levels?

6. Discuss the merits of supply-side policies aimed at reducing unemployment.

7. Explain how unemployment could be 'cyclical' or 'real wage'.

8. The government of a less developed country decides to reduce the extent of income and wealth inequality. What methods could the government use to achieve its goal?

9. How could knowledge as to the shape of the Phillips curve benefit governments? HL only

Extended response questions (25 marks each)

1. a) What are the main causes of unemployment? (10 marks)

b) How might governments lower the natural rate of unemployment? (15 marks)

2. a) Examine whether full employment is possible to attain (10 marks)

b) What possible social and economic costs might arise in trying to attain full employment? (15 marks)

3. a) What are the causes of inflation? (10 marks)

b) Explain how monetary and fiscal policies can be used to alleviate (= lessen) different types of inflation. (15 marks)

4. a) How could a government change its tax system in order to shift more of the overall tax burden from the poor to the rich? (10 marks)

b) Analyse why the government might wish to do this and the possible effects of the tax changes. (15 marks)

5. a) Is there a trade-off between inflation and unemployment? (10 marks) HL only

b) What are the implications of your answer in terms of economic policies aimed at growth? (15 marks)

Summary & revision

1. **Market-based supply-side policies** generally aim to help markets clear and increase allocative efficiency by removing market impediments.

2. **Labour market policies** include easier *hire-and-fire* legislation; cutting social and unemployment *benefits*; reducing *union power* and reducing/removing *minimum wages*.

3. **Tax incentives** on the supply-side involve lowering marginal tax rates on income tax; decreasing profit and capital gains taxes; and decreasing labour taxes.

4. **Critics** of supply-side policies point to high social costs of reducing social benefits; the time span it takes for supply-side policies to take effect; the limits of reduced taxes on income; the budget deficits that can result from cutting government spending; and clear evidence that unemployment is often demand-deficient in recessionary phases.

63. Free Trade

Key concepts:

- The gains from trade
- The World Trade Organisation

International issues are amongst the most controversial and divisive in economics, but I dare say that the majority of economists agree that trade and international exchange has been one of the most beneficial economic endeavours in humankind. Most economists view trade – more specifically *free* and *fair* trade – as one of the prime determinants of increased living standards in the world. Trade is good. Yet in spite of the benefits of trade having time and again proven to far outweigh costs, a good many interest groups will look upon my statement as highly political and therefore inappropriate to put into a school textbook.

However, my defence of the statement 'Trade is good' is not founded in a political standpoint but an economic standpoint whereas a good deal of the anti-trade rhetoric supporting statements such as 'Trade creates unemployment' is often erroneous and indeed sometimes nothing else but cheap political point-winning. A good deal of this section and the next will attempt to define 'good' in as measurably economic terms as possible. Then I leave it to you, the economist, to deliberate whether the statement is normative or positive – and right or wrong.

The core of international economics is comprised of issues which arise when independent nations interact in order to conduct economic transactions such as trade and foreign investment. The topics looked at here centre on the reasons for and benefits of trade; why protectionism exists; trade patterns and economic integration; the balance of payments; exchange rates; and the terms under which nations trade.

Ek Chuah, God of Trade, Tulum, Mexico. (circa 5th century AD)

TRADE

Why we trade

On the shelves around me here in my office/classroom as I write this sentence, one finds a bowl of bananas (Mexico), a thermos of coffee (Colombia/Kenya/Nicaragua), cigars (Cuba) and sugar (Denmark) – basically all the major food groups needed by writers. I also have a mobile phone, two computers, a hand-held computer, a USB memory stick and about 150 books on economics. With the exception of the bananas and my manta ray boots[1], very little has actually been produced here at home, yet here I am, happily eating a banana and sipping sweet coffee in anticipation of my evening cigar[2]. At the same time, my editor in Australia, Rory, is using silver from Mexican mines to frame his 'Economics Textbook of the Year' award. This must mean that someone in Tegucigalpa, Honduras, is walking to work wearing a pair of kangaroo skin boots – since my bananas didn't come from Australia but Honduras. In effect, the Mexicans sent silver to the Australians in exchange for boots, and used the Australian boots to trade for Honduran bananas. No, wait: the Aussies used Australian dollars (AUD) to buy Mexican pesos (MXN) with which they then bought silver; the Mexicans used the Australian dollars to buy Honduran Lempira (HNL) and then bananas; and the Hondurans exhanged the Lempira to get Australian dollars and R. M. Williams boots.

Now each country has a pile of foreign currency to be used at will. Mexicans on their way to The Great Barrier Reef in Australia can get some AUD at the airport in Toluca; a Honduran steel business can buy MXN at the bank in order to purchase steel rolling machines from ArcelorMittal in Mexico; and when Rory gets fed up with editing silly footnotes in economics textbooks he can buy HNL at the exchange office to import goods from Honduras and start a *Things Honduran* shop in Cunnamulla. This is an example of, respectively, trade in services, capital and goods. This silly example introduces five key reasons for trade;

1. Countries can trade *resources* which are not available at home.

2. Increased competition leads to lower *prices*, better *quality* goods and *innovation*.

3. Countries can specialise and achieve lower costs due to *economies of scale*.

4. Trading nations will be able to *consume outside their own PPF*. (See HL concept of comparative advantage in Chapter 64.)

5. Increased trade leads to *growth*, *development* and a possible *peace dividend*.

Factor Endowment – Swedish Ball Bearings and Honduran Bananas

The triangle of trade above shows a most basic economic fact that is all too often lost in the debate; the only way Mexicans can visit the Great Barrier Reef is by trading some silver. (Or rather, since most tourists don't produce silver themselves, they will use a domestic 'silver certificate' – MXN – which Australians will accept in order to buy silver and boots from Mexico.) Why do we trade instead of producing everything domestically?

Simple; growing bananas in my home country of Sweden would be only slightly more ludicrous than having herds of kangaroos in Kiruna in northern Sweden.[3] In a similar vein, the cigars on my desk are from Cuba but could just as easily have come from Nicaragua, Mexico, the Dominican Republic or 20 other countries. In fact, increasingly, my cigars *are* from Mexico and Honduras – but none of them are from Sweden, Norway or Finland.

You get the picture. Part of the answer lies in each country's **factor endowment** (= 'gift') in land – i.e. soil and climate – labour and capital. Honduras is perfect for banana cultivation and has abundant labour; Sweden has a 500-year history of producing industrial steel and invented the ball bearing; and Australia has kangaroos. The argument just put forward is easy to understand; each country has an abundance of certain factors of production which the others do not. This is the simple argument. A slightly more advanced argument is that the trading which takes place in different goods arises from a difference in the *relative* factor endowment of land in each country. Sweden could (and indeed *did*) grow tobacco and manufacture cigars, but would give up a

---

1 You think I'm joking, right? I'm not. People often ask me why I went to Mexico. Simple; it's one of the few places one gets to wear cowboy boots without going to a Halloween party. I have boots made from armadillo, ostrich, crocodile, lizard and manta ray. To general outrage amongst my students I told them I intend to put in an order for the USD1,000 *whale* skin boots. "Why?" they asked. "Well, I've got to get the order in before we run out of whales!" Eventually they figured out I was yanking their chains and didn't beat me up.

2 And, in all likelihood, a tequila. I've had about 15 bottles of booze on my classroom shelf in Mexico – presents from happy students and their parents. Yes, my boss knew. His smiling comment was "Matt, *everyone* knows. Just put a visible mark on the levels of the bottles so I don't see students staggering out of your classroom." Why am I so lucky with bosses and not wives?

3 But times they are a'changin'! On my way to work in southern Sweden a few years ago, I almost ran into an ostrich (!) on the road. It is apparently economically profitable in Sweden to farm ostriches domestically. Or is this globalisation – like the ostrich – run amuck? Whatever; get taller fences.

TRADE

great deal of resources in doing so, thus forgoing a great many SKF ball bearings which could be used to pay for an even higher quantity of Honduran cigars.

Countries are differently endowed with the factors labour and capital and will therefore have different cost ratios in the use of these factors; the higher the relative abundance of labour compared to capital, the lower the labour costs. Honduras has far more labour relative to capital while Sweden has the reverse situation. This is reflected in the domestic cost ratios for capital; Sweden is a most technologically advanced nation with very high labour costs while Honduras is a developing nation with low labour costs. The cost of using labour – relative to capital – will be much higher in Sweden than in Honduras. Honduras will therefore be far more likely to utilise labour in its production than capital and vice versa in Sweden.

Variety and Quality of Goods – Cars and Knives

Why is it then that so much trade takes place between countries with similar factor endowments?! The ball bearings in the previous example could just as soon be produced in any of the three which would mean that Swedish, Honduran and Australian ball bearings could be exported and co-exist on each of the countries' shelves. Germany and France share a common border and make thousands of similar products; France makes Sabatiér kitchen knives and Renault cars while Germany makes Zwilling Henkel kitchen knives and VW cars, for example. Why would France produce cars for export and then use the money to import cars from Germany? Resource endowment can hardly explain it and labour costs are higher in France than in Germany!

Consumer preference theory and **non-price theories of trade** (similar to non-price theories done by HL in Chapters 32 and 33) suggest that consumers desire *choice* and *diversity* of products, and the theory of non-price competition adds that consumers are willing to pay a premium for this benefit. Branded goods such as Adidas shoes (Germany), Toyota automobiles (Japan) and Nokia phones will be sold in countries which might well have domestic producers of the same goods, i.e. close substitutes. But a pair of Adidas is not the same as a pair of Brookes; a Toyota is not the same as an Audi; and a Nokia phone is not the same as a Sony-Ericsson. Firms continuously attempt to enhance their products' non-homogeneity via marketing, so such goods are all competing for our 'allegiance' to a brand and the intangible benefits for the consumer that go with that consumption. Whatever the underlying reasons, it is quite clear that most of the growth in trade deals with *intra-industry trade* (= trade in

same-industry goods), which now comprises between 60% and 70% of all trade in North America and in the EU.[4]

Gains from Specialisation - Economies of Scale

Trade is a transaction where two parties exchange goods and both feel better off – or as John Mill put it as early as in 1821; 'The benefit which is derived from exchanging one commodity for another, arises, in all cases, from the commodity *received*, not the commodity *given*.'[5] Two main features of trade that I try to put across to my people at an early stage are;

1. Trade has nothing to do with country borders

2. If a voluntary exchange takes place, then both parties have benefited – i.e. a 'win-win' situation rather than 'win-lose'.[6]

Trade is an economic activity that takes place between villagers in the village; between the villages in the countryside; and between the city and the countryside. I don't produce cigars or boots but I am an enthusiastic consumer of both. The way I am able to do this is by specialising in something I do reasonably well; warp young fragile minds with economics. The value placed on this earns me 'cigar/boot' certificates – which other producers ('specialists') accept as payment. Whether these producers are domestic or foreign is really beside the point.[7] Just view the economy as a firm or village, where tasks are allocated according to what each person does best and division

4 OECD; *The European Union's trade policies and their economic effects: Economics Department working papers no. 194, 1998*

5 Quoted from Globalisation – *Making sense of an integrating world*, page 74

6 One of my favourite in-class tricks in IB1 is to take off my watch – usually a rather pricey Rolex, Blancpain or Panerai – and "trade" it with someone wearing a cheap digital watch – lately my IB2 student Hardrock. "So, Hardrock, is there a winner and a loser?!" General howls of glee at the old man's loss and the student's gain. I then ask if anyone saw a gun or physical threat involved. Quieting down, the students then start to realize that if the transaction is free and voluntary we BOTH win! I seem to feel better off with the cheap watch and Hardrock feels better off with the Rolex. E.g. Hardrock has increased his total utility and so have I. Win-win. (There was almost some physical violence when he tried to sneak out after class with my watch on his wrist!)

7 I remember attending one of my first trade lectures at university where the professor continuously used the term 'international trade'. I found this rather tautological – similar to the usage of *Speiserestaurant* (literally 'Eating Restaurant') in German speaking countries. Actually, the professor was simply being most correct in his use of economic terminology since goods are in fact traded both domestically and internationally.

TRADE

of labour is utilised. Now expand this to include a neighbouring country. Adam Smith said it brilliantly in *Wealth of Nations*; 'The division of labour is limited only by the extent of the market'.[8] Drawing a line in the sand and sticking different flags on either side does not diminish the validity of Smith's statement. Trade gives products a larger arena of exposure and the possibility of being manufactured within *scale economies* while consumers enjoy a far greater range of goods.

'Impossible!'

Consumption Gains – Everybody is outside their own PPF!

Few, if any, of the present high income countries in the world would be where they are today without trade. It would be impossible for any single country to produce all the goods desired using domestic resources alone. Perhaps you are holding a ballpoint pen in your hand as you read this. Imagine the countless chemicals and materials that went into making such a simple (?) instrument; oil to produce plastic and dye, the many compounds in the ink, the chromium and steel for the nib (= point)....etc. Furthermore, consider all the machines involved in producing the raw materials ... and the machines to make the machines that produce raw materials .... I venture to claim that few economies could produce a simple ball-point pen based *entirely* on domestic resources. A single person would not live long enough to be able to produce the pen all alone. The solution is **specialisation and trade**.

Productivity Gains and Growth – More for all

An addendum to the *static* gains (of trade) in the above trade examples is the long-run issue of the effects of specialisation and economies of scale. It is fairly safe to assume that an

economy focused on a relatively narrow range of goods would over a period of time 'move up the learning curve', i.e. enjoy the benefits of increased *productivity* due to experience and innovations in production.

Even a country which has little in the way of domestic resources will benefit from specialisation due to ultimately hitting *economies of scale*, i.e. where the costs per unit (average costs) fall considerably as output increases. Just think of Japan and its specialisation in, amongst other areas, automobiles for the past 40 years. The production cost of each car has fallen in inflation adjusted terms throughout the time period. The Japanese have quite clearly shown that specialisation indeed increases productivity and lowers average costs over time.

Open markets, increased trade, increasing specialisation and benefits of scale all help to explain the astounding increase in the proportion of goods manufactured solely for export purposes. Figure 63.1 following illustrates how more and more of what is produced is subsequently sold on foreign markets; in the past 50 years world (real) GDP has increased by 500% while exports have increased by over 1,600% – over three times as much. During this period, the ratio of exports to GDP in the world rose from 7% to 15%[9].
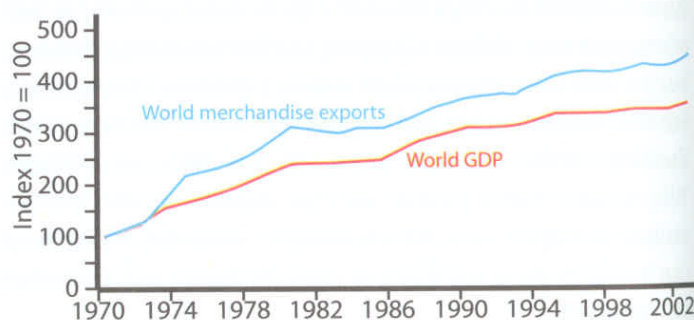

Figure 63.1 Trends in World merchandise exports and GDP, 1970 – 2003

(Source: Dept of Finance Canada at http://www. fin.gc.ca/ec2005/agenda/agc6-eng.asp)

The graphs in figure 63.1 have a remarkably powerful message attached, once you start to think about it. Any export must mean corresponding imports. As the increase in exports/imports far outstrips the increase in real GDP, then an ever-increasing proportion of world consumption expenditure – and therefore corresponding income – comes from trade. The powerful message is that the largest increase in world trade which the world has ever known coincides with the highest rate of increase in living standards in human history; the past 30 years. This makes it very difficult indeed to claim anything other than that trade adds to income and resulting prosperity and

8   Husted & Melvin, *International Economics*, page 57

9   *Globalisation*, page 72

welfare. If a country does well it will not 'harm' other countries it trades with – it will instead benefit them with more/cheaper/better goods. International trade is not a zero-sum game[10].

Political Gains - and a Peace Dividend

> "A day will come when we will see these two huge groups, the United States of America and the United States of Europe, standing face to face, clasping hands across the seas, exchanging their products, their business, their industry".
> Victor Hugo.

Just as Adam Smith claimed, integration and larger markets have meant economic growth. This is not only via benefits of scale, but by the transfer of technology, new ideas and production methods. Trade and the many associated activities are ultimately conducted by **people**; foreign workers, companies and ownership will all add to trading nations' fixed and labour capital as people exchange ideas and interact to become more efficient. In this process, there will naturally be clashes of cultures, norms and traditions, but also learning, understanding, acceptance and interdependence. People who are dependent on each other are highly unlikely to wage war against each other.[11] This is not a novel insight but something that nations have been aware of for centuries. Perhaps Guy Tozzoli, a famous trade and peace advocate who founded the World Trade Centers Association with members in over 100 countries, put it best: "When you're promoting business, you're promoting peace. Because when I understand your aims and your culture, I don't have any reason to declare war on you, and instead we work together. If we're doing business together, we're not going to fight each other because if I owe you money, you're not going to shoot me."[12]

Development – and lack lereof

An additional area where trade has proven highly beneficial to economies is within development. There is increasing evidence that open economies with increased trade has the highest rate of return of all development strategies. In 2003, Horst Köhler and James Wolfensohn (then Managing Director of the IMF and president of the World Bank, respectively) put it most explicitly; 'Expanding trade by collectively reducing barriers [to trade] is the most powerful tool that countries, working together, can deploy to reduce poverty and raise living standards.'[13] A paper from the IMF in 2010 puts in even starker terms: Lack of integration into the global economy is a major factor in the continued underdevelopment of the poorest countries.[14]

The World Trade Organisation

> "As the average age in the agricultural sector is so high (above 60) we are just waiting for them to die off." Chief Economist of the WTO, Patrick Low, when asked by one of my former students, Eric Blomquist, what the WTO was planning to do about the problem of agricultural subsidies and the CAP.

The WTO was born as the General Agreement on Tariffs and trade - GATT - in a hotel in **Bretton Woods**, New Hampshire in 1944, along with its 'sisters' the International Monetary Fund (IMF), the World Bank, and an exchange rate system subsequently known as the Bretton Woods system. The intention of the Bretton Woods conference was to create post-WWII systems and organisations which would lead to greater economic stability and increased trade around the world. Thus, in 1947, 23 nations signed the **General Agreement on Tariffs and Trade (GATT)**, which obligated members to commit themselves to trade under non-discrimination rules (the most favoured nation clause) and reduce tariffs on goods.

There were eight successive lengthy negotiations, called 'rounds', between 1947 and 1979, and by the seventh, the Tokyo Round, average tariffs had fallen from over 40% to 4.7%. The next round, the Uruguay Round, lasted from 1986 to 1993 and the 117 members expanded talks to include goods which had previously not been tabled for discussion, namely non-tariff barriers, agricultural/textile goods, investment, and intellectual property rights (e.g. patent and trademark rights). The Uruguay Round also replaced GATT with the **WTO on January 1st 1995**, giving the new organisation greater power and responsibility in international trade matters. The current round of negotiations started in 2001 in Doha, Qatar, and the 142 members agreed that it was time to put development firmly on the table. The

---

10 *Paul Krugman's Pop internationalism* is highly recommended reading on the lighter side of economics, where a number of most compelling arguments in support of win-win trade possibilities are given. I also recommend anything written by Naomi Klein for the other (frequently factually erroneous and dismally flawed) side of the story – *after* you've studied trade theory!

11 A trivial anecdote of the effects of economic integration is that there has never been a war between two countries that have McDonald's. Nor has there ever been a war between two democracies.

12 The New York Times, *Let There Be Peace, and Trade*, June 21 1998

13 Horst Köhler (Managing Director IMF) and James Wolfensohn (president World Bank), The Financial Times, Dec 12 2003, or see on-line version at http://www.imf.org/external/np/vc/2003/121003.htm

14 *Reaching the MDGs: An Action Plan for Trade*; IMF position note; September 16, 2010; SPN/10/14

Doha Development Agenda, as it became known, put the main emphasis on environmental and developmental issues in trade.

The WTO today[15]

The headquarters of the WTO is situated in Geneva, Switzerland. Some 600 administrators – bureaucrats – work there full-time for the now (May 2012) 153-member-strong organisation which accounts for over 95% of global trade. Over 30 additional countries are applying to join. It is headed by director-general Pacal Lamy from France, and had a budget of $US214 million in 2011.[16] All WTO agreements are both outlined and ratified (= signed) by member governments – thus far by consensus rather than majority vote. 600 administrators with a budget less than the 2,600 strong International Monetary Fund uses for travel expenses – hardly a threat to world prosperity, democracy and justice as numerous rather violent demonstrators have put it.

Aims of WTO

The function of the WTO is to create a global free trade environment by eliminating trade barriers and facilitating an arena in which to settle disputes. While GATT dealt primarily with eliminating visible trade barriers on goods, the WTO has increasingly cast a wider free-trade net, which now focuses on invisible trade barriers such as dumping and subsidies. Another aspect of the WTO's increased sphere of action is that not only goods are dealt with, but issues such as industry standards, labour rules and intellectual property. The WTO sets out five specific objectives in the multilateral trade system:

1. **Non-discrimination:** All agreements between members are based on treating other members equally, most prominently exemplified by the *most favoured nation clause, MFN* (see Chapters 74 and 75).[17] Any trade agreement for any good extended to one country must be granted for all members – exceptions are regional trade agreements such as NAFTA or the EU, but developing nations are allowed certain exceptions also.

2. **Freeing trade through negotiation/mediation:** The successive rounds focused on tariffs and quotas, but increasingly include non-tariff barriers and services. The WTO has considerably more power than GATT as an open market forum for multilateral negotiations and through its power as *mediator in trade conflicts* and its ability to enforce its verdicts.

3. **Stable trade environment:** The WTO aims to create a stable and predictable trade environment by setting *binding rules on maximum tariffs* and by demanding *openness and accountability* in trade policies in member countries.

4. **Fair competition:** The WTO does allow certain barriers such as anti-dumping tariffs and bans on hazardous goods, so while free trade is so far only a goal, the rules and regulations governing WTO members attempt to create a framework for a *level playing field* in terms of competition.

5. **Development:** It has been recognised by the WTO that while freer trade has often contributed to development, a good many of the poor member countries – about three quarters of WTO members – will face transitional difficulties in fully implementing reductions in trade barriers to designated levels. The poorest countries have thus been **granted extensions of current policies** and *special consideration*.

The five points above are a bit dry, but basically show how the WTO is based on the principles of **non-discrimination**, **equality** in rules and **fairness** in competition aimed at freeing trade between member states.

A Little Depth

The Doha Round – the Development Agenda (2001 – ?)

The ninth WTO Round kicked off in November 2001 in Doha, Qatar. The agenda was to continue along the lines outlined in the Uruguay Round, with particular emphasis on development

---

15 See excellent overview at http://www.wto.org

16 I simply must compliment the WTO for its incredible homepage. The site is clear, easy to use, has excellent sources – and is incredibly open in terms of "…who's paying the bill and what are they getting for their money…"

17 This is a bit like having a birthday party when you were a kid and you had to invite *all* your classmates rather than just the ones you liked. My mother tried this – once – when I was 7; "Matthew, if you don't invite Johnny, there will be no party!" There was no party. I just didn't like him. (The following year we did have a party – with a more discriminating guest list.)

– thus the name **Doha Development Agenda**. The development issues listed for negotiation primarily outlined the obligations of developed countries to continue to lower – and in some cases eradicate – all forms of trade barriers on agricultural goods and textiles outlined in the Uruguay Round. A successful completion of the Round would mean some $US300 billion in increased income in poor countries, pulling over 140 million people out of poverty, according to the World Bank.[18] The Doha Round was to be concluded by December 31st 2004.

To be blunt, the Doha Round has thus far been a massive failure. During the Ministerial Conference (the higher level negotiations of WTO members) at **Cancun**, Mexico, in September 2003 the talks broke down and it seemed that the distance between developing and developed countries was wider than ever. Staunch EU, US and Japanese resistance to tariff and export subsidy reductions was met by the formation of an alliance of 21 developing nations – G21 – which further polarised the positions of countries supposedly negotiating. The meeting ended when the chairman prematurely closed negotiations, and delegates have since then pretty much blamed everyone else for the failure.

A less heated but no more successful outcome came from the **Geneva** negotiations in July 2008. One of the key issues was a proposal for a Special Safeguard Mechanism (SSM), originally proposed by the so-called G33 – a group of developing countries cooperating in trade issues. The SSM allows developing countries to impose special tariffs on key agricultural goods under certain conditions – e.g. unexpected decreases in world prices and huge increases in imports. This was the pivotal point of disagreement and arose between India which had the SSM as a key demand, and the US (a large exporter of agricultural goods) which opposed the SSM. It is highly unlikely that the Round will be completed during 2013.

Summary & revision

1. **Reasons for trade** include:

    a.   Counties are differently endowed with resources

    b.   Trade leads to increased competition and thus better quality and lower prices

    c.   Industries in trading nations can attain economies of scale

    d.   Specialisation allows nations to consume outside their respective PPFs

    e.   Trade has been shown to be pro-growth and pro-development.

2. The World Trade Organisation (WTO) is a multilateral and membership-based organisation which aims to lower barriers to trade and provide a forum for solving trade disputes.

TRADE

---

18 *The Economist, The WTO under fire*, Sep 18th 2003

64. HL extension – Absolute and Comparative Advantage

Key concepts: HL extensions

- Absolute advantage
- Comparative advantage
- Assumptions and PPFs
- Opportunity cost ratios and possible terms of trade
- The gains from trade – the CPF vs the PPF for each economy
- Calculating opportunity costs using data
- Evaluation of comparative advantage

One of the legends floating around amongst economists is that Nobel laureate Paul Samuelson was once challenged by the sceptical mathematician Stanislaw Ulam to come up with one single proposition or theory in economics which was not either *trivial* or *impossible to prove* in reality. Samuelson went away, brooded, and came back with his reply; comparative advantage. "That it is logically true need not be argued before a mathematician; that it is not trivial is attested by the thousands of important and intelligent men who have never been able to grasp the doctrine for themselves or to believe it after it was explained to them."[1]

Nations trade simply because it is more economically efficient to have some goods made abroad and then trade for them. If code for computer programs is increasingly being done in India by American firms rather than in the US, then it simply means that America is getting it done cheaper than domestically. American consumers gain and India benefits from higher incomes. America will ultimately have to export other things in order to be able to import, since no one will sell goods to the US unless they find something worthy of their US dollars.

Absolute advantage

When a country can produce goods at a lower cost, i.e. using fewer resources, than another country it has an absolute advantage. Another way of putting it is that a country has an absolute advantage if the unit of output per unit of input is higher than another country's. A simple example using two countries, Mexico and the USA, and two goods, textiles and agricultural goods gives us Figure 64.1:

1 *Defending free trade*, Arvind Panagaria, *Economic Times* November 22, 2000

Potential output of the two countries assuming factor inputs are equal in quantity				
	Agriculture	Textiles	Consumption in autarky (= no trade)	Consumption in specialisation and trade
USA	8	4	4A and 2T	4A and 4T
Mexico	4	8	2A and 4T	4A and 4T
'World' totals			6A and 6T	

Figure 64.1 Absolute advantage in one good

USA's resources (= factor endowment) enables it to produce 8 units of agricultural goods or 4 units of textiles – in any combination. With the same quantity of resources Mexico can produce 4 or 8 units respectively. Clearly the US is more efficient in the production of agricultural goods and Mexico is more efficient in textiles – the US has an absolute advantage in agriculture and Mexico an absolute advantage in textiles. Assuming autarky (= self-sufficiency and no trade) and each economy allocates half of the resources for each good, total 'world' output and consumption is 6 units of agricultural goods and 6 units of textiles.

Now assume the US and Mexico open up their economies to trade, specialise 100% in their respective absolute advantage, and trade half of the output with the other economy on the basis of 1 unit of agricultural good for 1 unit of textile. The US is producing 8 units of agricultural goods and exporting

4 units – which of course means that US imports 4 units of textiles. Mexico produces 8 units of textiles and imports 4 units of agricultural goods. In summa; both economies are now *consuming* outside the limits of their own production.

Most of those who insist on opposing trade as a method of increasing overall wellbeing make a classic beginner's mistake by confusing *absolute* advantage with *comparative* advantage, commonly phrased along the lines of 'Well, but if a country is disadvantaged in *both goods* then it cannot benefit from trade'. This is a very common statement, but most erroneous as shall be seen.

Comparative advantage[2]

The concept of absolute advantage is quite straightforward and understandable at an almost intuitive level: Imagine that you and a classmate are given the task of doing a 15 minute presentation in economics in front of the class. Odds are you would divide the labour so that each one does what he/she is best at, say illustrations on overheads or writing the script.

The more challenging issue arises in explaining how both can still benefit from trading if one of you happens to be better at *both* tasks. This basically means finding where the other is *least disadvantaged* – and still divide the labour. In other words, the one who is worse at both illustrations and writing will at least have a *comparative* advantage in one of the two. This simple example is understood by every brother and sister who have been given a set of chores to do together, and for all the seeming complexity of the coming example it still boils down to the basic issue of the 'weaker' partner doing the tasks in which he/she is least disadvantaged.[3]
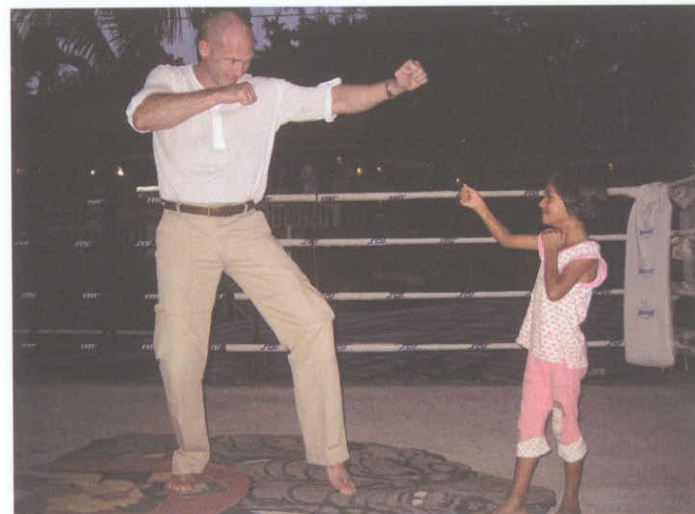
Uh oh…looks like I'm both weaker *and* slower… where's my comparative advantage?!

The lesson here is that an economy with an *absolute advantage in both goods* might still benefit from trade if there is a comparative advantage. The theory of comparative advantage was set down in the early 1800s by the famous English economist **David Ricardo**, who understood that, while the concept of absolute advantage was very powerful in its argument for free trade, it was too limited to have general applicability and acceptability in the real world. In reality many countries are better than others in the production of any number of goods, so Ricardo set forth to show that even one country having an absolute disadvantage in many goods could be mutually beneficial to both countries when trade takes place.

In our example using the US and Mexico producing agriculture and textiles, we now assume that the US has an absolute advantage in production of *both* goods and can produce either 10 units of agricultural goods or 10 units of textiles. Mexico can produce 4 units of agricultural goods or 8 units of textiles. Mexico in fact actually gives up *fewer* resources in the production of textiles than the US does – a unit of textiles 'costs' 1 unit of foregone agricultural goods in the US while the 'cost' is only 0.5 units of foregone agricultural goods when Mexico produces textiles. In core economic terms, Mexico's opportunity costs in the production of textiles are lower than the US'. Mexico therefore has a *comparative advantage* in the production of textiles.

TRADE

---

2   The last thing I wish to do is confuse you – but since it is on the list, here goes. The year **1048** was … nothing special really. Pope Damasus II dies, Emperor Henry III names his cousin Bruno the new pope Leo and the Chinese emperor Shenzong is born. This has *absolutely nothing* to do with economics…except for the number **1048**.

3   All my students get to hear about fighting in the karate finals. If your opponent is stronger than you but slower, you fight him/her on speed – where you have an absolute advantage. If he/she is faster but weaker, your absolute advantage is in strength. If he/she is 50% stronger and 10% faster…well, you have a problem. However, I recommend fighting him/her on speed where you are *least disadvantaged* – where you have a comparative advantage – and you will thus get the least beating.

Definition: 'Absolute and comparative advantage'

If the US can produce more agricultural goods and textiles than Mexico using the same amount of resources, than the US has an **absolute advantage**.

If Mexico has a lower opportunity cost (in terms of foregone production of agricultural goods) in the production of 1 unit of textiles, then Mexico has a **comparative advantage** in the production of textiles.

Here is the focus of this example: 1) Set down relevant *assumptions* and outline *production possibilities* for two countries; 2) establish *cost ratios* and *terms of trade*, i.e. an exchange rate; 3) and finally, we use the cost ratios to show how specialisation and trade will allow each country to *consume outside its PPF*.

Assumptions and PPFs for the US and Mexico

So, we assume that there are only two economies, USA and Mexico, and two goods, agricultural goods and textiles. Again we assume that economies can attain maximum potential output – Pareto optimum – and move along the production possibility frontier, but that no economies of scale exist. We will also assume that opportunity costs are constant – this is not absolutely necessary but it makes the example easier to follow. Finally, we assume that neither trade barriers nor transport costs exist. Diagrams I – III in Figure 64.2 show the situation facing the two economies:

- Initially there is no trade between them, whereby the US produces/consumes 6 units of agricultural goods and 4 units of textiles (point **A** in Diagram I)

- Mexico produces/consumes 2 units of agricultural goods and 4 units of textiles (point **B** in Diagram II).

- Diagram III shows each economy's PPF in relation to the other. Notice how the *PPFs are divergent*, which is central to your understanding of comparative advantage.
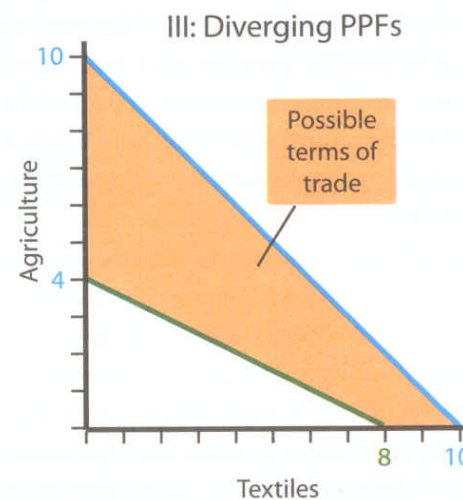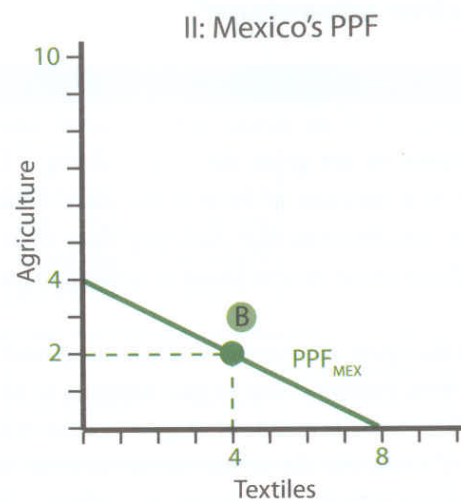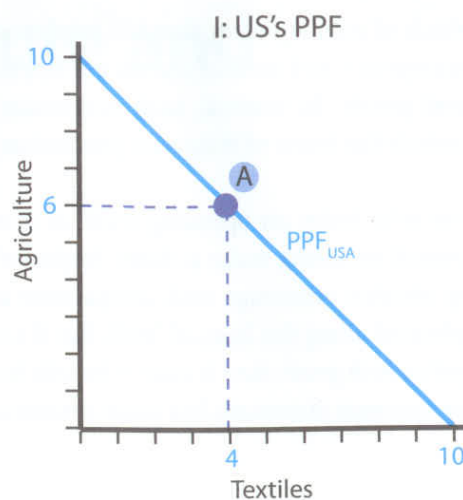
Figure 64.2 (I – III) PPFs and cost ratios for US and Mexico

Opportunity cost ratios and possible terms of trade

The theory of comparative advantage states that, *ceteris paribus*, if a country has a comparative advantage then there are gains to be made in trade. Are there comparative advantages shown in diagram III in Figure 64.2? Go back and re-read the definitions of absolute and comparative advantage and apply them to Figure 64.3.

	Potential output		Opportunity costs (in terms of the other good; producing 1 unit of agricultural goods in Mexico means foregoing 2 units of textiles)	
	Agricultural goods	Textiles	Agricultural goods	Textiles
USA	10	10	1	1
Mexico	4	8	2	0.5

Figure 64.3 Opportunity cost ratios for agricultural goods and textiles

Clearly the answer is "yes", since Mexico – which is absolutely disadvantaged since the US can produce more of both goods – has a lower opportunity cost in the production of textiles than the US. Mexico foregoes 0.5 units of agricultural goods in producing 1 unit of textiles whereas the US foregoes 1 unit of agricultural goods in producing a unit of textiles. Mexico has a comparative advantage in textiles.

I: Cost ratios of 1 unit of textiles

II: Cost ratios of 1 unit of agricultural goods
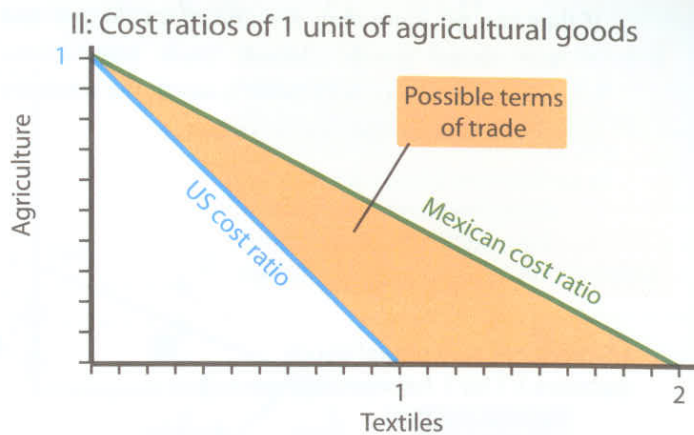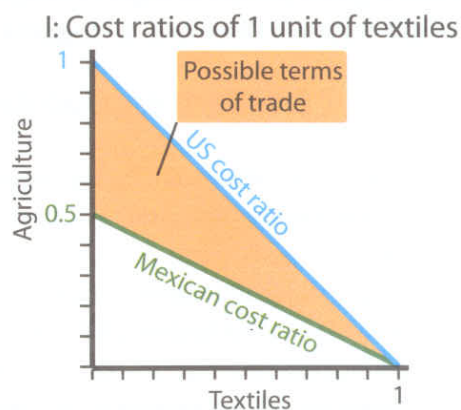
Figure 64.4 Opportunity cost ratios and terms of trade

The **terms of trade** (Figure 64.4, diagrams I and II) are shown as the sandy-coloured areas in-between the diverging cost ratios.

- **Mexico** can trade domestically at a rate of 1 unit of textiles for 0.5 units of agricultural goods and would therefore be willing to trade with the US if the return – imports! – were *more than 0.5 units* of imported agricultural goods for each unit of exported textiles.

- The **US** in turn pays 1 unit of agricultural goods for each unit of textiles produced domestically, and would trade if it *paid less than 1 unit* of exported agricultural goods for each unit of imported textiles.

- The **terms of trade** will lie somewhere between 1 unit of textiles trade for between 0.5 and 1 unit of agricultural goods (diagram I) or, alternatively, 1 unit of agricultural goods trades for between 1 and 2 units of textiles (diagram II).

-

The gains from trade – the CPF vs the PPF for each economy

Let's assume that **both economies specialise 100%**, and then look at the issue of trade, terms of trade and the consequent gains from trade. (I shall look at these issues from the point of view of Mexico in order to stress that even the country with an absolute disadvantage in both goods can gain from specialisation and trade.)

- **The US** now has 10 units of agricultural goods which could have been traded domestically at the ratio of 1 unit of agricultural goods for each unit of textiles. The trade incentive is that the Americans will trade

if they can get *more than one unit of textiles* per unit of agricultural goods. (When 'trade negotiations' take place the object is to make it worth the while for citizens of Mexico; they too must gain.)

- **Mexico**, having specialised 100% in textiles, has 8 units of textiles where each unit can be traded domestically for 0.5 units of agricultural goods, i.e. each unit of agricultural goods costs Mexico (as in opportunity costs) 2 units of textiles. If Mexcio can pay less than 2 units of textiles for each unit of agricultural goods there will be an incentive to trade. Figure 64.5 shows each economy's menu of alternatives and preferences.

US produces agricultural goods (receives 1 T for 1 A domestically)

Mexico produces textiles (pays 2 T for 1 A domestically)

Agric.	Textiles				Agric.	Textiles
1	1.0	USA's limit!	Mexico's limit!		1	2.0
1	1.1				1	1.9
1	1.2				...	...
1	1.25				1	1.25
...	...				1	1.2
1	1.9	Mexico's			1	1.1
1	2.0	limit!	USA's limit!		1	1.0
1	2.1				1	0.9
↓	↓				↓	↓

Better ToT for US!

Better ToT for Mexico!

Terms of trade established at
- 1 unit agricultural goods = 1.25 units textiles
  Or:
- 1 unit textiles = 0.8 units agricultural goods
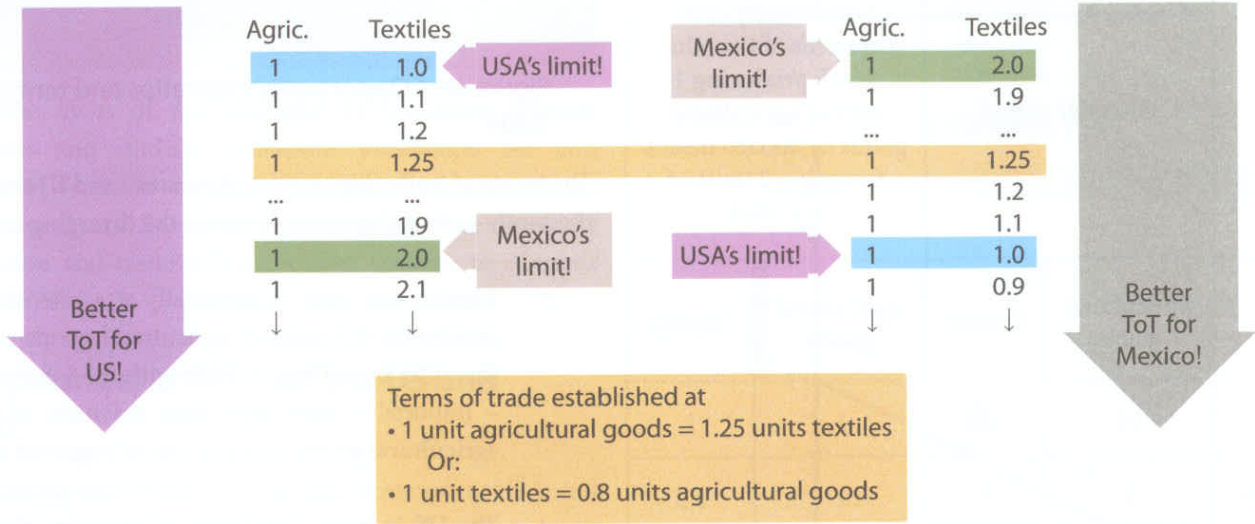
Figure 64.5 Possible terms of trade for US and Mexico

The US is completely specialised in agriculture, with domestic trade possible at 1:1 for textiles. If the terms under which the US can trade agricultural goods for imported textiles – **the terms of trade** – are better than this ratio, trade will take place:

- The more textiles the US receives for each unit of agricultural goods, the better the terms of trade are for the US. This is shown in the blue column/arrow on the left in Figure 64.5, where increasing quantities of imported textiles per exported units of agricultural goods represent improving terms of trade for the US:

- Mexico, totally specialised in textiles, will have the reverse situation (expressed in payment per units of agricultural goods from the US), and will not give up more than 2 units of textiles for 1 unit of agricultural goods. Mexico's terms of trade will improve if it can pay, i.e. export, less than 2 units of textiles in payment for each unit of imported agricultural goods. The green column/arrow on the right show Mexico's improving terms of trade.

Let us finally assume that the two countries arrive at an **equitable trade ratio** (= terms of trade) of **1 unit of agricultural goods to 1.25 units of textiles.** (Alternatively; consider 1 unit of textiles for 0.8 units of agricultural goods.)[4] Assuming furthermore that Mexico wishes to remain at 4 units of textiles in domestic consumption, it will be able to export the remaining 4 units of textiles, receiving 3.2 units of imported agricultural goods. The US, having specialised in agricultural goods, exports these 3.2 units of agricultural goods in order to import 4 units of textiles. The results of this trade are illustrated in diagrams I and II, Figure 64.6.

---

4   Quite naturally, terms of trade are not 'negotiated' or 'arrived at'. Terms of trade **evolve** over time in accordance with each country's demand for imports (i.e. relative preferences) and supply/price of domestic goods.
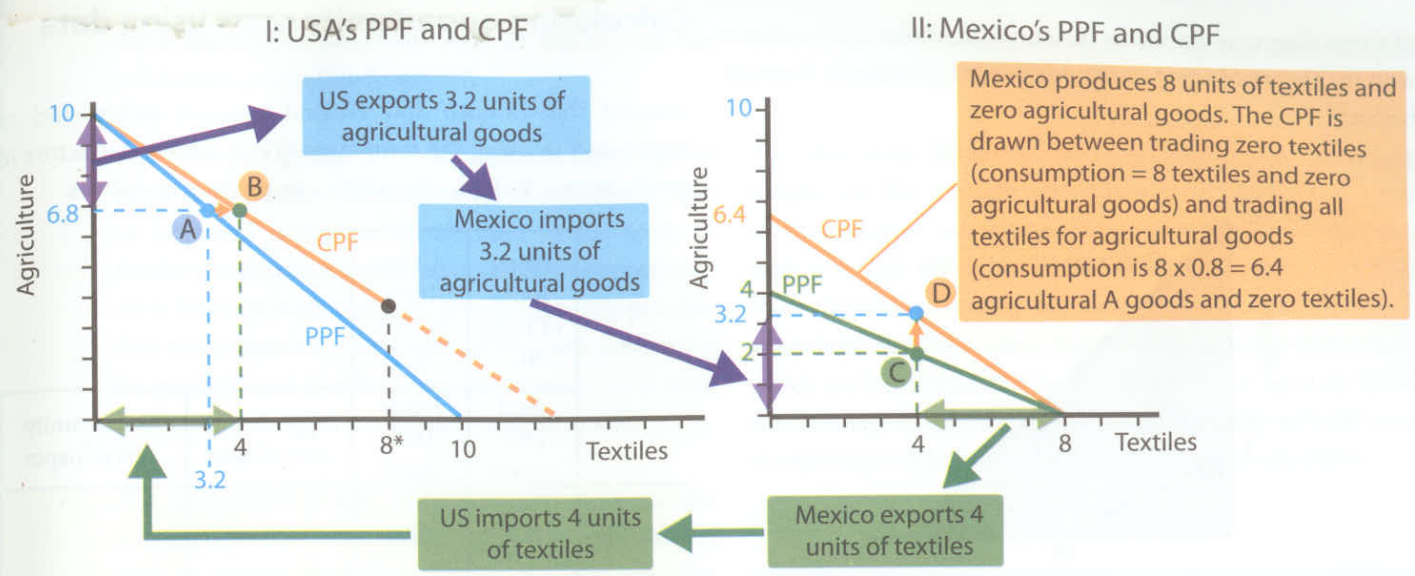
TRADE

Figure 64.6 PPFs and CPFs for the US and Mexico

POP QUIZ 64.1

1. Why is the CPF for the US dotted at textile consumption levels above 8 units?

2. What would US imports be if they instead exported 5 units of agricultural goods?

3. What would happen to the US and Mexican terms of trade if the Mexican CPF were to 'swivel upward', say going from 8T⇔6.4A (as in the example) to 8T⇔8A? (See Figure 64.4 for help.)

Win-win outside the PPFs

The example shows how differing opportunity cost ratios between countries create comparative advantages. When these advantages are used by reallocating resources and specialising, both countries can consume outside their production possibility frontiers while total output increases. Figure 64.6 shows that total 'world' (in our example consisting of the US and Mexico) production and consumption has increased from 8 units of agricultural goods and 8 units of textiles to, respectively, 10 and 8 units. Herein lies the argumentative power of the theory of comparative advantage; **more is produced using the same**

amount of resources and both parties can increase total consumption.[5]

Cost Ratios Yielding a Comparative Advantage

The key point in respect to trade is to realise that comparative advantage arises only when countries' **PPFs are divergent**, i.e. when the *opportunity cost ratios differ* between countries. If cost ratios in producing the two goods are identical in country A and country B, there can be no gain for either country in specialising and trading – at least not in terms of quantity. Figure 64.7, diagrams I to III shows three possibilities between Portugal and Spain in the production of wine and wool. Which

---

5 It is important to understand the very powerful *principle* of comparative advantage. I once had a fierce argument with a rather left-leading colleagues who was adamant that one should pay a maid exactly the same as *I* would demand for doing the same service. This is of course utter woofle ('woolly' plus 'waffle' – the former means mental confusion and the latter means hazy and unclear) since my opportunity costs in ironing shirts is much higher than my maid's opportunity costs. I am good at ironing actually, and cooking, and sewing (it's the kind of thing men who can't keep wives learn over time), but the fact remains that for me to spend 20 hours a week doing house chores would mean forgoing quite a few hours of rather productive economics. A specialised maid can do the job in less time and for lower opportunity costs than I. The money I earn in 20 hours of writing/lecturing can pay for her opportunity costs several times over.

of these diagrams shows an incentive for trade in accordance with the theory of comparative advantage? (Answer in footnote below.[6])
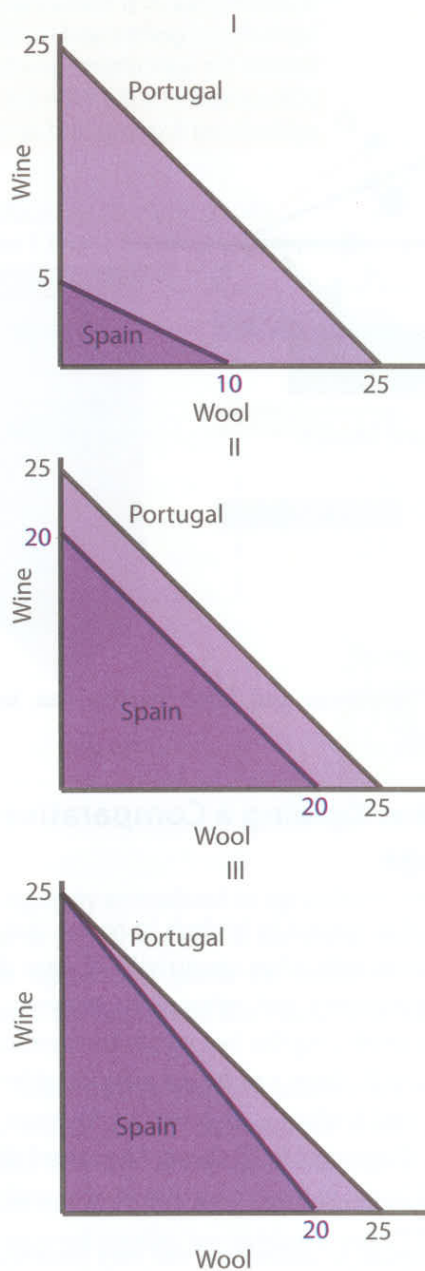
Figure 64.7 Possible PPFs for Portugal and Spain

Calculating opportunity costs using data

Assume that Sweden and Finland produce either steel or paper. and in using the same quantity of resources (factors of production = 'FoPs' in the table), can produce as follows:

	Using 100% of FoPs in...	Using 100% of FoPs in...		
	Steel (mn units)	Paper (mn units)	Opportunity cost of steel	Opportunity cost of paper
Sweden	800	1000		
Finland	320	800		

Assume that both nations are Pareto efficient, that no barriers to trade exist, the goods are homogeneous, no economies of scale are attainable, and that factors of production are 100% mobile.

1. Which nation has an absolute advantage?

2. Is there an incentive for trade according to comparative advantage?

3. If so, which nation will specialise in and export steel?

4. What are the possible terms of trade? Draw a cost ratio diagram for support.

5. Assume that each country specialises 100% and that the nation specialising in paper uses half of output domestically and exports the rest. Assume valid terms of trade and draw each country's CPF based on your terms of trade 'exchange rate'.

(If you get stuck, have a look at point 4 in the end of chapter Summary and revision.)

Evaluation of comparative advantage

Does the theory of comparative advantage stand up to the test of reality? Well, yes, albeit a "yes" with some notable qualifications (= conditions):

- Quite evidently no country will **specialise completely** as in our model, which lowers the possible specialisation gains and thus the amount available for trade. This will naturally limit the extent to which

---

6 **Answer;** diagrams I and III show that quantitative gains in consumption are possible since the opportunity cost ratios for Spain and Portugal are different. Diagram II is not conductive to trade according to the theory of comparative advantage since each country has identical opportunity cost ratios and neither can benefit (quantitatively) by trading with the other.

the consumption possibility frontier lies outside the production possibility frontier.

- The theory of comparative advantage is based on the **premise of product homogeneity**, which, in the light of increased non-price competition in the trade of similar goods, is not explained by different opportunity costs between trade partners. It is theoretically possible that two countries with identical domestic cost ratios for goods X and Y still trade since consumers in both countries wish to have a greater range of goods.

- The assumption of **zero transport costs is unrealistic** – especially for bulky goods with low value-added such as empty plastic containers and iron ore. For such goods the proportion of transport costs to total production costs will be very high, which might negate any comparative advantages of trade.

- Another assumption subject to criticism is that there are **no barriers to trade**. This is perhaps one of the strongest prevalent forces against the comparative advantages of countries.

- The model used here is based on **constant opportunity costs**, which are highly unlikely in the real world. However, as pointed out earlier, constant costs were assumed to make the example easier to follow and a "convex" production possibility frontier could in fact have been used.

- The theory also neglects to incorporate the significant **benefits of scale** that have been shown to exist, thus failing to explain the very large increase in trade between similar economies, primarily OECD countries. However, as shown earlier in Chapter 24, economies of scale *support* our conclusion that specialisation leads to lower average costs, which in fact strengthens the argument of international specialisation benefits.

- Finally, the theory of comparative advantage does not deal with **income distribution** within countries. Trade will create both winners and losers in terms of domestic income; sectors competing with relatively cheap imports will see incomes fall while export sectors will show increased income. Whether the winners' gains outweigh the losers' losses is a most complex question, dealing once again with the vagaries of fairness. Most economists would prefer to view income distribution aspects and trade benefits as separate issues and look

at distributional effects in the light of domestic policy rather than trade policy.

The objections listed can in some cases carry considerable weight, yet the strength of the now over 200-year-old theory of comparative advantage lies primarily in illuminating a *basic principle* which has been empirically shown to hold true; trade depends primarily on comparative rather than absolute advantage. Studies confirm that it is often the case that countries which are less productive in all industries (i.e. have an absolute disadvantage in all areas) than another country will still be able to export goods in which they are relatively competitive.[7]

POP QUIZ 64.2

Reasons for Trade

1. Assume two neighbours; Maria is a professional house painter and Johann is an auto repair man. Using concepts from this section, explain why it would be beneficial for Maria to paint Johann's house while he repairs her car.

2. Look up Jamaica and Iceland in a World Factbook and then use trade theory to explain why Jamaica is one of the world's largest exporters of bauxite and Iceland is a major producer of aluminium. Use the same reasoning to explain how Icelanders can consume far more bananas than are imported! (Eh, the aluminium issue is not related to the banana issue.)

3. Explain why a worker at the Nissan car plant in Japan might drive a Ford.

4. Country X and Y both produce steel and plastic. Country X has an opportunity cost ratio of 1 unit of steel to 1 unit of plastic; country Y has an opportunity cost of 1 unit of plastic to 0.5 units of steel. Using the theory of comparative advantage, explain the trade possibilities.

5. Continuation of Question 5: Assume instead that country Y has opportunity costs of 0.5 units of plastic to 0.5 units of steel. What are the possible terms of trade for country Y?

TRADE

---

7 See for example *International Economics*, Krugman & Obstfeld, pages 23 – 34

6. Why do economists generally regard the theory of *comparative* advantage as a very strong argument in favour of trade?

7. Can a country benefit from trade even though it has higher production costs per unit of output in all goods than other countries?

8. How might a country improve its comparative advantage?

9. How might exporters view a worsening of the terms of trade? How about consumers?

Summary & revision

1. When an economy can produce more than another using equivalent quantities of factors of production, one says an **absolute advantage** exists.

2. A country that has a lower opportunity cost in producing a good (in terms of foregone output in other goods when factors of production are re-allocated) is said to have a **comparative advantage**.

3. In exemplifying how an economy can benefit from trade due to a comparative advantage, we make a number of **assumptions**:

   a. Only two goods and two economies

   b. No barriers to trade or transport costs

   c. Goods are homogeneous

   d. No economies of scale

   e. Constant opportunity costs

   f. Both nations are Pareto efficient

   g. No transport costs

   h. 100% specialisation for each economy

4. **Opportunity cost ratios** show the cost for each economy in the production of goods. Only when opportunity costs differ can the theory of comparative advantage be used to explain why trade would take place.

5. The **terms of trade** lie in between the cost ratios. In the figure above, the terms of trade will lie between 1 unit of paper for 0.4 to 0.8 units of steel. Alternatively, one unit of steel for 1.25 to 2 units of paper.

6. Trade along the lines set out by the theory of comparative advantage enable economies to consume outside their respective PPFs. Each economy will have a **consumption possibility frontier** (CPF) that 'swivels' outwards from the point where the PPF intercepts the axis showing 100% specialisation in a good. The better the terms of trade, the more the CPF lies outside the PPF.

7. There are some notable **limits to the theory of comparative advantage**:

   a. Products are *not homogeneous*

   b. *Barriers to trade* do exist and skew comparative advantage towards the protectionist economy

   c. *Transport costs* can negate comparative advantage

   d. *Economies of scale* exist – yet this would in fact *increase* the gains of an economy specialising in certain sectors

65. Restrictions on Free Trade

Key concepts:

- Tariffs
- Quotas
- Subsidies
- Administrative barriers
- Health and safety regulations
- Environmental standards
- Other forms of protectionism (some very interesting ones!)
- Infant industry argument
- Anti-dumping barriers
- Other forms of protectionism (some very interesting ones!)
- Reasons for protectionism
    - Protection of industries and employment
    - Infant industry barriers
    - Anti-dumping barriers
    - Strategic arguments
    - Source of government revenue
    - Means to overcome a balance of payments disequilibrium
- The case against protectionism
    - Inefficiency in resource allocation
    - Comfort zones of protectionism
    - Increased prices of goods and services
    - Effect on domestic competitiveness

"When goods can't cross borders, armies will."
Frédéric Bastiat

Trade has been shown both historically and theoretically to have multiple and resounding economic and social benefits. Global trade is growing at roughly twice the pace of global output – which means that more and more of what we produce is exported and what we consume is imported. In spite of this, there are many barriers to trade between countries arising from various protectionist policies. It is worth commenting on the fact that protectionism is almost solely the result of *government* policies (sometimes under pressure from domestic industries) and not due to decisions made by *people* buying the goods. Most protectionist measures are also against the expressed recommendations of economists.

The main reason for this lies in the seeming inability of economists and various interest/lobby groups to see eye-to-eye. Economists estimate the societal value of trade through the looking glass of the 'benefits of consumption' which in essence views *increased imports as the benefits of exports*; more exports

facilitate increased imports.[1] A more common view amongst interest/lobby groups, ostensibly acting on behalf of industries/labourers, is that *exports are the gain and imports the costs*; more exports generate more income and employment.

This section attempts to address the main reasons for limiting trade and whether the costs of this outweigh free trade. HL students are recommended to revise the incidences of taxes/subsidies, resource allocation in context with consumer/supplier surplus, and long-run average costs.

TRADE

---

1   Keynes put it brilliantly, as usual; "Imports are receipts and exports are payments. How as a nation can we expect to better ourselves by diminishing our receipts? Is there anything a tariff can do, which an earthquake could not do better?" Taken from Husted & Melvin, page 201

What is 'free trade'?

Totally free trade does not exist anywhere, let us set that straight from the outset. Military goods, plutonium and coca leaves will be subject to rigorous trade limitations for example. Even mainstream goods and services such as textiles or fruit will be subjected to a number of more or less obvious trade barriers – in fact, it seems that Hong Kong is the only country in the world which has no taxes whatsoever on imports.

Free trade is when countries can exchange goods and services unhindered by any form of market intervention which disadvantages a country in selling its goods on a foreign market. Anything which increases the costs or reduces market accessibility for foreign firms relative to domestic firms constitutes a barrier to trade. In short, only when a foreign producer can sell goods in another market under the same conditions as their domestic rivals, can one speak of completely free trade.

Definition: 'Free trade'

Free trade is the total absence of any form of intrusion or barrier in the flow of goods and services between countries which disadvantages foreign firms in favour of domestic firms. Free trade means that the only additional disadvantages experienced by a firm on a foreign market arise due to transaction costs such as foreign exchange costs, transportation costs and poorer market knowledge and closeness to foreign markets. Protectionism can be widely regarded as any form of action taken by a country whereby the price of domestic goods become more favourable relative to imported goods.

Protectionism is commonly divided into *tariff* and *non-tariff barriers*, where the former is basically a tax on imports and the latter is any of dozens of other possible barriers to trade, notable examples of which are quotas, subsidies to domestic producers and various forms of regulatory rules on imported goods.

Tariffs

When the European Union (EU) puts a 140% tax on imported sugar – notoriously one of the most protected goods in the EU – but not domestic sugar, it has imposed a **tariff**. A tariff can be an *ad valorem* tariff based on the import price as the sugar example, or a *specific* (unit) tariff, e.g. the $US 2.36 on a bottle of wine entering the US.[2] While tariffs are by far the most common form of trade barrier, the average tariff rates in the world have fallen drastically in the past 70 years, from over 40% in the 1940s to less than 4% today. This doesn't really mean that there is less protectionism today but rather that highly visible trade barriers such as tariffs have to a great extent been replaced by less obvious barriers.

Definition: 'Tariff'

A **tariff** is a tax levied on imported goods. The tariff can be both an ad valorem tax, e.g. a percentage of the import price, or a specific tax, e.g. a tax based on a measurable unit such as tonnes or bottles.

From Autarky to Free Trade to Tariff

Assume an economy which goes through three stages; an initial stage of zero trade followed by perfectly free trade and finally a tariff. Figure 65.1, Diagrams I to III, illustrate what would happen to a fictitious economy producing and consuming oranges.

I: Closed economy

TRADE

2   United States Dept of Agriculture at: http://www.fas.usda.gov/info/agexporter/1997/August%201997/uswine2.html

Figure 65.1 Autarky, free trade and tariff

horizontal world supply curve S_{world}. This open economy will now have domestic consumption at Q_{D_1} and domestic output at Q_{S_1}. The difference is made up of imports of $Q_{S_1}óQ_{D_1}$. This is the **free trade equilibrium**.

- Now, if a tariff of $P_1èP_2$ is levied on imported oranges, shown by the arrow in diagram III, the world supply curve shifts from S_{world} to $S_{world\ +\ tariff}$, increasing the price on the domestic market from P_1 to P_2. This will have the effect of encouraging/enabling increased domestic production in a manner similar to that of a minimum price (see Chapter 15). Domestic producers replace $Q_{S_1}óQ_{S_2}$ tonnes of imported oranges with domestic, which shows how the tariff protects domestic producers by raising the price of imported oranges. The increase in the domestic price of oranges also has the effect of decreasing the quantity demanded from Q_{D_1} to Q_{D_2}. The combination of a higher price level and lower demand achieves the purpose of lowering imports and raising domestic production.

Effects of a Tariff

There are other issues involved in the imposition of tariffs. Figure 65.2 shows – now using fictitious figures for clarity – how the tariff will have both **transfer effects** on society and also **costs to society** in the form of waste. Domestic suppliers are able to increase their output from 500 tonnes of oranges to 600 tonnes while imports fall from 400 tonnes to 200 tonnes. Foreign producers' revenue diminishes to area E after having lost revenue in areas F and G. The grey trapezium of A+B+C+D shows how consumers lose by paying a higher price for less quantity consumed; consumers could be consuming 900 tonnes of oranges at a price of €1 per kg but are instead paying €1.2 per kg for 800 tonnes per year. (You will recognise this area as the gross loss of **consumer surplus**).

In Figure 65.1, diagram I, the country is consuming only what it produces, i.e. it is a closed economy and produces/consumes Q_0 tonnes of oranges at P_0 per tonne.

- When the economy opens its doors to foreign competition, diagram II, it will face a horizontal world supply curve (we drop the assumption of "only two countries in the world" so world supply will to all intents and purposes be infinite) of S_{world}. The world price of oranges, P_1, is lower than the domestic price of P_0 leading consumers to increase their quantity demanded to Q_{D_1} at a market equilibrium of S_{world} = D. (Note that we are assuming that domestic and foreign oranges are perfectly substitutable.) Observe that the effect of an open economy on *domestic output* is negative; S_{dom} shows that domestic producers are competitive up to Q_{S_1}, whereafter foreign producers are able to undercut domestic suppliers along the

Figure 65.2 The effects of a tariff on oranges

Not all of the area A+B+C+D in Figure 65.2 is a net loss to society since parts of it represent a transfer to other parts of society. Another childish example: Picture that you lose your wallet in the forest and it contains €100; the wind blows open the wallet and spreads €90 to other happy members of the forest community while €10 is never found. Society as a whole has lost only the €10 since your loss of €100 has been offset by others' gain.

In our trade example, the domestic economy is made up of three players; consumers, domestic suppliers and government.

- **Area A**: A loss to consumers which is offset by a gain to one of the other two players means that society is in fact equally well off. Thus area A shows how domestic firms benefit from selling more at a higher price: the increase in *supplier surplus*.

- **Area C**: Area C is the gain to government from additional *tax receipts* of the tariff times imports; 200 tonnes at 20 cents per kg = €40,000.

- **Areas B and D**: Areas B and D represent the net reduction in overall economic welfare; the **net societal loss**.

- **Area B** shows the cost to society of production being taken over by relatively inefficient producers; at all levels of output between 500 and 600 tonnes, foreign producers could have produced oranges at a lower cost. The triangle shows the extra resources used by society in producing domestically rather than importing from more efficient foreign producers at the world price. This can be labelled a 'green' loss since it is wasteful to allocate productive resources to areas best left to others.

- **Area D** shows the *net* loss of consumer surplus, i.e. the loss of consumer surplus which has not been transferred to others in society (domestic firms and/or government).

Taken together, areas B and D constitute a pure allocative waste to society called a **deadweight loss** (which you will recognise from Chapter 8).

The loss of consumer benefits and the allocative losses arising due to the imposition of a tariff are among the strongest

arguments in favour of free trade. In effect, tariffs serve to penalise domestic consumers in favour of domestic producers – in effect subsidising more inefficient production by transferring additional income from consumers to domestic firms. However, there are possible offsetting gains to the domestic economy arising from tariffs and general protectionism, such as safety, environmental and employment issues. We will look into some of these issues further on.

Quotas

We now turn to other forms of trade barriers, i.e. *non-tariff barriers*. A common method to protect the domestic economy is by setting actual quantitative limits on the amount of imports, a **quota**.

This form of non-tariff barrier has the effect of raising the domestic price of the good just as a tariff, but has different transfer effects – the wallet contents are found by different forest citizens. Figure 65.3 shows how the domestic economy imposes a limit on imports of oranges rather than imposing a tariff. (To make the comparison with tariffs easier, assume that a quota of 200 tonnes is set.) When foreign suppliers are confronted with a quota that halves the quantity they are allowed to sell to another country, simple supply and demand analysis suggests that the highest possible price will be charged. The maximum price foreign suppliers can charge for the 200 tonnes of oranges is €1.2. Since foreign suppliers raise the price in order to receive the highest possible price on their goods, the market price is forced upwards and domestic suppliers increase supply from 500 to 600 tonnes per year.

> **Definition: 'Import quotas'**
>
> A **quota** is a quantitative (physical) limit on imports set by the importing country. The effect of a quota is that domestic suppliers will replace some of the imports at a higher than world market price. Foreign suppliers will have an incentive to raise the price of their goods, thus making a windfall gain.
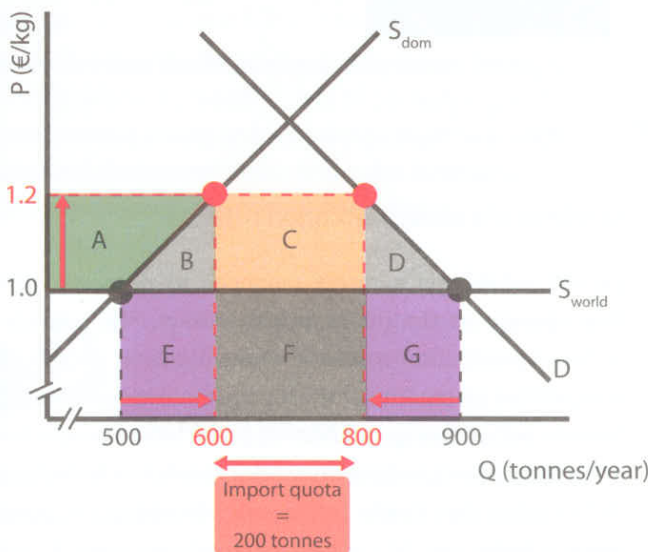
Figure 65.3 The effects of a quota on oranges

A+B+C+D; loss of consumer surplus

A; gain of domestic supplier surplus

B; efficiency loss I = "Green loss" to society

C; windfall gain to foreign suppliers (€40,000)

D; efficiency loss II = net loss of consumer surplus

E & G; foreign suppliers' loss of revenue (€200,000)

F; part of original revenue of foreign suppliers (€200,000)

TRADE

Again there are societal transfers and net efficiency losses. Figure 65.3 above outlines these as previously demonstrated with tariffs. The main difference between tariffs and quotas is within the issue of which societal members will benefit from the difference between the import price of €1 and the new price of €1.2 resulting from protectionism. **Area C** does not go to government tax coffers as in the case of tariffs, but into the pockets of foreign suppliers. This unexpected increase in revenue is known as a **windfall gain**, and partially offsets the loss of revenue shown by areas F and G. (Another difference is that an increase in domestic demand will have completely different effects on domestic output and foreign imports. Under tariff protection, an increase in domestic demand will be met by an increase in imports whereas under a quota system the increase in demand will induce domestic suppliers to fill the increased demand.)

Subsidies

Any action by government which lowers the ratio of domestic prices to import prices (in other words *raises the relative price of imports* or *lowers the relative price of domestic goods*) will serve as a barrier to trade. Therefore, when domestic producers receive direct payments for production or *indirect* production incentives such as tax breaks or low-interest loans, the effect is that of a **subsidy**. A subsidy lowers the production costs (HL: marginal costs) for domestic producers and enables them to be more competitive with foreign producers. Please refer back to Chapters 14 and 15 for a little depth on the 'hidden subsidy' debate and the Common Agricultural policy.
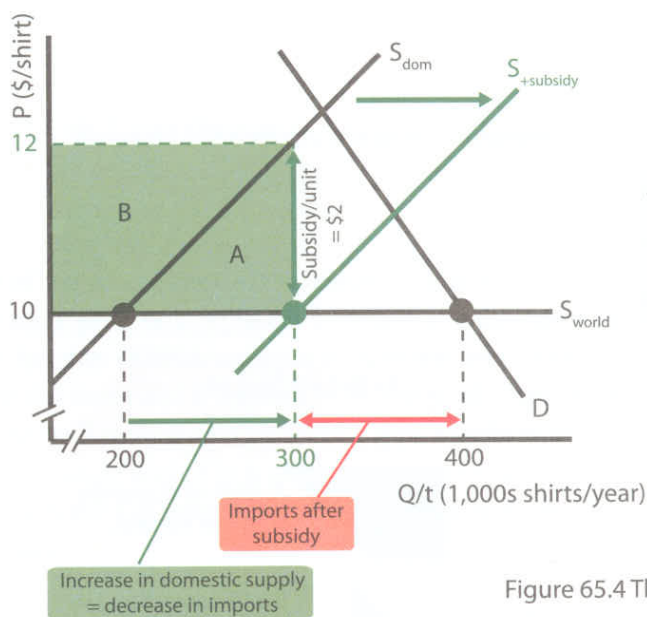
Efficiency loss – the additional units produced domestically (100,000 shirts) all have higher production costs (HL: MC) than S_{world}. This is suboptimal allocation of resources.

Total cost of subsidy to government; $600,000 ($2 x 300,000) per year

Figure 65.4 The effects of a subsidy on cotton T-shirts

Figure 65.4 shows how a subsidy going to domestic producers affects the market for cotton T-shirts in an economy. Initially the domestic economy is producing 200,000 T-shirts and importing 200,000 T-shirts, enabling domestic consumption of 400,000 T-shirts at a price of $10. When the government subsidises domestic production at $2 per T-shirt, domestic suppliers are able to increase supply from 200,000 to 300,000. This is shown by the shift of the domestic supply curve from S_{dom} to $S_{+subsidy}$. (Note that the entire incidence of the subsidy goes to domestic suppliers and that there is still an efficiency loss triangle.) The subsidy allows domestic producers to capture market share from foreign producers who are now left with one quarter of the market rather than half.

Subsidies can take other forms, two of the more noticeable being export subsidies and guaranteed price schemes. An export subsidy is a direct payment to domestic producers that operate on the international market. The subsidy serves to lower production costs and make domestic producers more competitive internationally. This mechanism enables domestic firms to sell on foreign markets at prices below production costs – which is known as dumping and is absolutely forbidden under WTO rules. (See below under Anti-dumping.) A **guaranteed price scheme** can be considered a subsidy since a minimum price is created and the excess supply can be 'dumped' on the international market via export subsidies.

Administrative barriers

When my mother and father moved to Austria in the late '80s they made the mistake of bringing a foreign car with them. It took more time and money to comply with Austrian rules on

headlights, colour of blinkers, ground clearance etc than was worth it.[3] This is but one example of national legislation which in fact acts to create – unintentionally for the most part – barriers to trade. When foreign auto producers have to spend time forcing through *administrative red tape* the delay adds to overall costs.

Administrative barriers take on many forms, such as the restriction of foreign ownership; licensing rules; excessive documentation on place of origin; extensive testing of new products (very common for pharmaceuticals…see 'Other barriers…' further on) and translation of all documentation into the language of the importing country are but a few examples.

It is also quite possible that *international developments* increase existing barriers or create new ones, such as more stringent security controls at all points of entrance to a country such as airports and shipping ports. Increasingly, in these times of worries over international terrorism, the International Maritime Organisation (IMO – a UN agency) has increased security demands on many ports around the world. Instituting such safeguards as fenced off and guarded port areas, security codes and baggage checks for all personnel and visitors to the port and security cameras will inevitably add to delays in shipping and increased costs for shipped goods.[4]

> **Definition: 'Administrative obstacles as barriers to trade'**
> **Administrative trade barriers** arise when foreign producers face increased costs in complying with administrative issues in importing countries, such as content documentation and sales licences.

However …

It is notably easy to implement intentional administrative barriers; a classic example of such was when France decreed in 1982 that all imported (e.g. Japanese) videocassette recorders had to clear customs at a tiny customs house in Poitiers – far from any and all major ports and points of entry to the country. This backlogged the imported VCRs (not 'Very Cunning Restrictions' but 'Video Cassette Recorders') tremendously and limited the imports to a trickle – giving domestic producers time to 'catch up'.

Health and safety regulations

An extension of administrative barriers is when domestic legislation decrees that imported products must meet certain **health and safety criteria**. When I was writing a final paper at university, I visited a large Swedish knife manufacturer, Krång Johan Ericsson, which sold knives all over the world. One of the managers told of how the company was in the final stages of getting approval from the US Food and Drug Administration (US FDA) for a series of ergonomic butcher knives for American slaughterhouses. The company had already spent over two years and shipped countless prototypes of the knives to the US to be subjected to successive rounds of testing prior to approval for use in the food industry. The process was very costly for the knife firm and the only problem was apparently that the handles contained certain chemicals necessary to give the knife handles a good grip and the FDA demanded with absolute certainty that there would be no leakage into the meat.

> **Definition: 'Health and safety standards as trade barriers'**
> Health and safety standards in importing countries can cause both increased costs of product modification and delays for foreign producers in getting the goods to market. This advantages domestic producers.

There are literally thousands of sets of standards which countries impose on products in the interest of public health and safety, which all serve to increase costs to importers and/or limit the physical access to the market. Here are a few examples of such standards: Moped manufacturers in the EU will face varying speed limits and the engines will have to be accordingly modified to reduce top speeds; legislation on content labelling will vary greatly in different countries and force manufacturers to print a variety of different labels; certain food additives, colouring and preservatives will be outlawed in some countries; bans on advertising goods such as alcohol and tobacco are increasingly common; and fireworks will face numerous safety regulations as to size and to whom they may be sold.[5]

---

3   No, I'm not making this up; the *blinkers* were the wrong colour! I've always wondered if this might have something to do with colour blind Austrian cows crossing alpine roads.

4   *Sydsvenska Dagbladet;* '*Säkerheten I hamnarna höjs*', Friday 12 March 2004

5   And bans – admittedly a *very* good idea in Mexico.

TRADE

Misuse of health and safety concerns?

Perhaps the most (in-) famous example of health standards which limit trade is the EU ban on all beef containing growth hormones which was implemented in 1989. This has had the effect of a de facto ban on American beef, since virtually all US cattle are treated with growth hormones. Successive American administrations have all claimed that the EU policy is clearly in violation of WTO rules of free trade, since no conclusive evidence exists that growth hormones in any way endanger the health of beef eaters. The WTO, which is the mediator in such disputes, ruled in favour of the US in 2003 yet the EU has as yet not complied. The Americans retaliated with tariffs on $US100 million worth of EU agricultural goods and the ensuing tit-for-tat retaliatory 'tariff game' between the worlds' two major traders came close to escalating to a trade war.

In March, 2008, the WTO again criticised the EU ban on beef from hormone-treated cattle as scientifically unjustified. US Trade Representative Susan Schwab applauded the WTO re-statement; "The findings confirm the principle that measures imposed for health reasons must be based on science."[6] Another, perhaps equally famous, example of the EU – US trade conflict is the EU restriction on imports of genetically modified (GM) produce such as soybeans and corn. This issue, like that of beef, has yet to be resolved.

OUTSIDE THE BOX

Spurious (= fake) health and safety concerns

It is historically clear that a number of regulations ostensibly part of a health and safety code can be considered **covert** (= hidden, concealed) **trade barriers**. The *Reinheitsgebot* – purity laws – in Germany regulated what could be used to make beer and had the effect of banning most imports. This restricted

6   The UN Mission to the EU – see http://useu.usmission.gov/ Dossiers/Beef_Hormones/Mar3108_WTO_Dispute_Panel. asp

access was lifted in 1990 in line with the increasingly integrated market of the EU.

Italy had a similar experience when previously banned imported pasta – made from the wrong kind of wheat – gained access to the Italian market at the same time. France has banned imported barbeques as 'unsafe'; Belgium has laws demanding extensive labelling on imported foodstuffs; Japan previously 'inspected' imported tulip bulbs by slicing them open (!) and still has a safety inspection on every single imported car – at a cost to the importer of several thousand US dollars; and Canada takes an average of twice the amount of time to approve new US pharmaceuticals than the notoriously tough US FDA[7].

My personal favourite regulation which in fact was a covert trade barrier was the EU ruling on camel cheese imported from Mauritania; the EU Commission ruled that while the low cholesterol/ high protein cheese was indeed healthy, basic sanitary requirements dictated that the cheese could only be imported if the camels were milked by machine. What the EU ruling failed (?) to take into consideration was that the camels were owned and milked by *desert nomads*[8]. Let me get this straight; the camels should carry electrical milking machines around in the desert? Talk about the final straw…

Environmental standards

As trade growth continues unabated, it is increasingly common for countries to demand that imported goods meet with domestic **environmental standards**. For example, the US – with California at the forefront – has implemented a number of laws governing emission standards and fuel consumption on cars, which any and all imported cars have to conform to. This forces non-US auto manufacturers to adapt engine performance to fit the American market, adding to costs. Another example was the ban on aluminium cans and certain non-reusable plastic containers in Denmark during the 1980s in favour of re-usable glass bottles which were considered environmentally friendly.[9] A number of European producers of soft drinks and beer lodged numerous complaints with the European Commission (the

7   See the official EU site at http://europa.eu.int/comm/internal_ market/en/goods/infr/602.htm and the Independent Canadian public policy institute, The Fraser Institute at: http://oldfraser. lexi.net/publications/forum/2002/02/section_06.html

8   Businessworld, *Don't say Cheese*, December 30, 2002

9   A number of studies show in fact that aluminium cans have less environmental impact than glass bottles since cans use less energy in (re-)production, cleaning and transportation than glass bottles.

administrative power of the EU) since this posed a trade barrier on their exports to Denmark. Their case was strengthened by the fact that Danish producers continued to export beverages in cans to other EU countries, and the ban has since been rescinded by the Danish government.

Definition: 'Health and safety standards as trade barriers'

Health and safety standards in importing countries can cause both increased costs of product modification and delays for foreign producers in getting the goods to market. This advantages domestic producers.

Increasingly in our ever-integrating world, environmental standards are set not so much to meet domestic demands but to set global standards for environmentally acceptable production. This has arisen in partial response to increasing developed nation outsourcing of production to developing countries where environmental standards can be noticeably lacking. This is the '**race to the bottom**' argument, which argues that negative externalities such as pollution, diminishing forests and encroachment of wildlife are 'exported', in a manner of speaking, to poorer countries in order that firms in rich countries might avoid increased costs of strict domestic environmental policies. To countermand this, a number of countries have implemented environmental standards which limit – or ban – imports of goods which do not meet certain criteria.

The debate about environmental standards in trade

The issue of using environmental arguments in trade is highly contentious (= controversial), as it is incredibly difficult to distinguish between commendable environmentalism and contrived trade barriers. The WTO has a number of articles which specifically allow for measures which are 'necessary to protect human, animal or plant life or health'...*provided* the measures do not constitute "an arbitrary or unjustifiable discrimination between countries where the same conditions prevail or a disguised restriction on international trade".[10] For example, the US imposed a ban on tuna fish imports caught using methods known to kill a large number of dolphins and another ban on shrimp from several Asian countries that used nets which also trapped turtles; the WTO ruled both bans illegal to general outrage amongst environmental groups.

Other environmental arguments are much less contentious, such as the restrictions on imports of certain hardwoods from rainforests in many OECD countries, where a range of certification procedures are in place to see to it that goods made of hardwood come from plantations rather than rainforests. Perhaps one of the most well-known examples of how environmental standards have been beneficial is that virtually all developed countries have long since banned goods either containing or made using chlorofluorocarbons (CFCs).[11] Taken as a whole, many administrative, health/safety and environmental standards are sometimes referred to as covert trade barriers. This does not mean that the barriers are necessarily intentional but that differing standards disadvantage foreign producers. Yet it bears repeating that the line between legitimate and contrived barriers of these types is blurred and subject to continuous debate and international litigation.

Other forms of protectionism (some very interesting ones!)

If you think I was ranting above, baby you ain't seen nothing yet![12] While my poor editor is desperately trying to get me to cut down on pages, I simply cannot avoid mentioning a few additional forms of protectionism. It's what economics authors do for entertainment.[13]

Currencies and exchange rates

Currencies and exchange rates: Exchange rate fluctuations will create uncertainty for exporters and importers, leading perhaps to less trade. Commissions (service charges) on foreign exchange transactions also create barriers. Many countries also have strict controls on the amount of currency which can be traded – and by whom. A country can also devalue its currency (i.e. lower the exchange rate, which is the 'price' of the domestic currency) thereby making imports more expensive and domestic goods cheaper relative to imports.

---

10 *McCulloch et al, Trade liberalisation and poverty: A handbook*, page 327 and Article XX of GATT 1994

11 *Legrain, Open world – the truth about globalisation*, pages 241 – 253

12 Bachman Turner Overdrive, 1974

13 We are pretty sad creatures.

Government favouritism in procurement and nationalism

Government favouritism in procurement: Government agencies will frequently be required to favour domestic industries. Look closely at the brand of cars you see used for official functions in countries which have domestic auto production.

Nationalism: Persuasive advertising slogans and campaigns such as 'Buy American/French/Italian…' can skew demand away from imports which might be both better and cheaper.

Pirates of the ~~Carib~~…that's Malacca in Indonesia

I thought the Mexicans were masters of protectionist policies… and then I moved to Indonesia. The Indonesians make the Mexican trade barriers look like something Donald Duck invented when Goofy got bored. Here are three incredible examples of trade barriers erected by the Indonesian government.[14]

Cabotage laws … or 'sabotage' laws?[15]

In the early 2000s, Indonesian ship owners lobbied fiercely for government protection against foreign owned transport vessels operating in Indonesian waters. Foreign cargo ships carried about half the goods transported in Indonesia. Why didn't more Indonesian ships compete with the foreign ships you say? Well, it seems that the Indonesian ship *builders* had previously campaigned hard to get protection against foreign ship builders…which raised the price of ships…so Indonesian shipping companies *couldn't afford to buy more ships*. It gets better! The Indonesian governments response was, yes, you guessed it; *more* protectionism. This time in legislation passed during 2008 which basically will ban foreign owned vessels from transporting goods within Indonesia's 18,000 islands. I'm not done yet! To make sure that foreign ships don't offload in ports where there are logistical benefits in getting Indonesian flagged carriers to the final Indonesian destination, a new law

(2012) limits all foreign ships to just two ports.[16]

Maybe the Indonesians should ask the Hawaiians and Alaskans what they think. The Jones Act from 1920 does exactly what the Indonesian government has done; all goods and services to and from Hawaii and Alaska from any other US states must be transported on US flagged ships. Guess which are two the most expensive states in the USA? I'll give you two guesses but you'll only need one.[17]

Self medication

Along with a decree in 2010 forcing importers to renew their import licenses – an effective short run ban for many of the companies due to the horrendous inefficiency and corruption of Indonesian customs – the government took special pains to limit the importation of pharmaceuticals. Yes, you read correctly. One of the poorest nations in the world with some 120 million people living in poverty, has passed legislation whereby any foreign company making medicines must open a manufacturing plant in Indonesia within two years.

When the health minister was asked to comment on the legislation his response was "If they [the foreign companies] want go away, go ahead."[18] This in a country where fake drugs constitute around 25% of all drugs sold and which the government admits poses a serious threat to public health.[19] Who do you think this hits the hardest, the low income groups or the high? Well, let's just say that everyone I know in Indonesia goes to Singapore for any type of medical service. Do you think my maid can send her kids to Singapore to get cheap generic drugs?

"Pay me or I shoot the package!"

The bald-faced corruption in Indonesia involving any type of goods shipped into the country is legendary – and a bit scary. I actually took the time to get hold of the complete list of tariff rules translated into English. Scary reading! It turns out that *any* good coming into the country can be subject to import

---

14 In writing these examples I am well aware of the risk of being kicked out of the country or jailed for slanderous disrespect towards the Indonesian government. If you don't see a fourth edition, you'll know where I am.

15 I'd love to pronounce this 'Sabotage' in this case but it's pronounced 'kabotage'. It means legislation regulating the transportation of goods and people within a country's borders – not international transports.

16 http://online.wsj.com/article/SB1000142405274870371250457 6232231094375232.html

17 Hawaii at number one and Alaska at number two.

18 US Chamber of Commerce; http://www.uschamber.com/ sites/default/files/international/asia/southeastasia/files/ dcindonesiaoped.pdf

19 See the World Health Organisation at http://www.who.int/ medicines/services/counterfeit/impact/ImpactF_S/en/index1. html and the Office of the US Trade Representative at http:// www.ustr.gov/sites/default/files/uploads/reports/2009/NTE/ asset_upload_file255_15479.pdf

TRADE

tax if the good is valued at over USD50. But if you read the fine print, it is also the case that no matter what it says on the shipping label in terms of declared value, the customs officials in charge *get to assess the value themselves!*

Basically goods are held hostage at the post office or customs office at the port of entry. Why not read them the law? Well, one colleague tried just that. He got a translator and a copy of the law and asked the duty officer for an explanation and an official paper outlining why he was supposed to pay a USD100 'tax' for his own (used) computer he had sent from his address abroad. He got no explanation…but he did get his laptop. It had a big boot print on it and was smashed to pieces. My lady Bell had the same thing happen – for an old iPod she forgot in Australia and had sent to our home here. They wanted USD45 for 'duties' – but forgot to get paid before delivery to our door! I cannot put in print my letter to the customs officer in charge. Finally, the winner; Katie at school is an IB examiner and was presented with a USD40 'customs invoice' for a pack of…wait for it…IB exam papers! Since there is zero fiscal value of exam papers and the declared value was exactly that, the cunning bandits put the tax on the *weight* of the package!!! When our fabulous Indonesian admin staff called them and challenged them about the fee, they said "Oh never mind!" and dropped the issue. No, I am not making this up. I mean, you can't *make* this stuff up.

Reasons for protectionism

Standard economic theory states that protectionism generally creates losers amongst foreign import firms and domestic consumers; winners amongst domestic firms; possible gains to government in the form of tax receipts; and overall allocative losses for society. Thus far, most economists' conclusion would be that costs outweigh the gains of protectionism, which merits asking why there are so many avoidable barriers to trade still in place. A follow-up question must address the issue of whether there are in fact trade barriers which carry net societal gains, i.e. where the benefits are greater than the costs. While there are no clear cut answers to any of these questions, a number of arguments can be put forward in defence of protectionist measures. There is strong disparity amongst economists on the degree of validity of virtually all these arguments, so in some cases I will include a *pro and con* addendum.

Protection of industries and employment

The employment protection argument is probably the most controversial argument for protection. It is certainly the most divisive amongst economists, where views range from 'partially viable' to 'absolutely erroneous'. The basic argument in implementing trade barriers in reference to unemployment is that traditional/large industries which are subjected to increased foreign competition will cause *structural unemployment*. When comparative advantage ratios shift as foreign competitors become increasingly proficient in certain industries, governments often resort to protectionist measures in order to preserve jobs – or at least delay a decline in what becomes a 'sunset industry'. A large proportion of the trade barriers of the past 20 years fall within this argument; steel, cars and cotton in the US; sugar, wine and ship building in the EU; and agricultural goods and textiles – to some extent – in all developed countries.[20]

The argument often enjoys broad support amongst both citizens and government officials, since cheap foreign imports and concomitant loss of jobs is an easy connection to make and the perceived societal benefits are clear. Anyone who has seen the effects of massive layoffs in a staple industry, such as in Northern England during the mine closures in the early 1980s will easily understand the appeal of softening the blow by restricting foreign competition.[21] Increasingly the argument is extended to limit 'the exportation of jobs' to low cost labour countries resulting from the increased ease with which firms can move capital. When firms such as Flextronics (a component manufacturer for mobile phones) shut down entire factories in the EU and moved production to China in order to take advantage of lower labour costs, the resulting sectoral unemployment inflicted severe social and economic costs on the job exporting country. In light of this, it would appear that protectionism is indeed warranted.

Pro and con in employment argument

The employment argument has *limited validity* at best and is *seriously* flawed and fallacious at worst:

Pro: The argument can have *limited validity* when protectionism is used to grant domestic industries *temporary* relief from debilitating (= devastating) foreign competition in order to grant time for domestic product and factor markets to clear and reallocate factors to other industries. This reasoning is often applied in the case of important domestic industries that are in a state of permanent decline, so-called *sunset industries*. Protectionist policies might also be complemented by various

20 Again, Hong Kong being the exception.
21 Recommended movies: *Brassed Off*, *The Full Monty* and *Billy Elliot*.

labour market policies such as retraining and education to facilitate a smoother transition period. However, such policies have a tendency to become entrenched and therefore long run, leading to numerous negative effects, primarily higher prices of goods and efficiency/competitive losses.

Con: Economists claiming that the infant industry argument is *seriously flawed* commonly put forward the following three points:

1. Recall that in order to import one must export. No country will be able to buy imports 'on credit' in the long run. When the world's number two cotton producer, the US, protects domestic cotton growers from imports from the number one cotton producer, China, the effect is to limit Chinese ability to buy Caterpillar bulldozers – increasing unemployment in this other US industry. This is a classic example of the **fallacy of composition** in economic argumentation; what is good for one industry is not necessarily good for *all* industries. Protecting cotton jobs in the US thus comes with a triple price tag: 1) The protectionism is often in the form of subsidies which cost taxpayers via government funds; 2) Textile firms pay a higher price for cotton and consumers pay a higher price for cotton goods; 3) And jobs lost in road machinery manufacturing inflict additional economic and social costs.

    Official US statistics show overwhelming evidence that US trade with China is indeed not a one way street: exports of iron, steel and copper from the US to China increased by over 100% from 2002 to 2003. During the period 1997 to 2003, Chinese exports to the US increased from $US65.8 billion to $US163.3 billion, which is an increase of 148%. During the same period, US exports to China went from $US12.8 billion to $US28.4 billion, an increase of 122%.[22]

2. Protectionism might save jobs in domestic production but *jobs will be lost in the import sector*, where people are employed in wholesaling, retailing and servicing imported goods. During the past 20 years in the US, during the most trade-expansive and open period in US history, 40 million *more* jobs were created than lost; strong indicative evidence supporting the economic truism that trade creates rather than destroys jobs.[23]

3. Protecting the domestic economy from cheap foreign labour is perhaps the most erroneous and misleading argument of them all and is almost always advanced by rich countries in reference to poor countries.[24] The main fallacy of this argument is the assumption of a '**lump sum**' (= fixed amount) of available employment opportunities and that cheap foreign imports remove employment opportunities for domestic labourers.[25] In fact, the opposite has continuously shown to be true, where increased openness, trade and competition has been met in rich countries with rising productivity and thus higher wages. While this evolution has indeed created periods of labour redundancies and unemployment, higher real wages have triggered increased demand for a multitude of other products over the past 20 to 30 years. This has primarily increased the demand for **services**, many of which cannot be imported and therefore are supplied domestically. So while manufacturing jobs in the US fell from just under 21 million to under 20 million between 1970 and 2000, over 50 million new jobs were created in services during the same period.[26]

Infant industry barriers

The infant industry argument was put forward almost 200 years ago and is thus one of the longest standing arguments in favour of protectionism. Suppose that a country has potential comparative advantage in the production of a good, but is competing on an international market where global benefits of scale exists and the market is mainly comprised by entrenched

---

Fund and James Wolfensohn, President of The World Bank; *We can trade up to a better world*, Financial Times, December 12, 2003

24 I hate to say it, but even Albert Einstein fell for it. Quote: "*If two factories produce the same sort of goods, other things being equal, that factory will be able to produce them more cheaply which employs fewer workmen- i.e., makes the individual worker work as long and as hard as human nature permits. From this it follows inevitably that, with methods of production as they are today, only a portion of the available labor can be used.*"

25 An anecdote herein is the story of a trade economist visiting China in the 1970s and seeing how hundreds of road workers were using shovels. He suggested that a few bulldozers and other road machines would do the job far more efficiently. His horrified host replied that this would create enormous unemployment. The economist subsequently suggested that if the aim was employment creation the shovels might be replaced with spoons. (Taken from *Defending Free Trade, Arvind Panagariya*; *Economic Times*, November 22, 2000)

26 As a percentage of employment, manufacturing jobs have gone from over 26% to under 15% during the 30 year period, while services increased their share to 72% in 2000. See US Bureau of Labour Statistics at www.bls.gov/fis/fls/flslforc.pdf

---

22 US International Trade Commission, US Department of Commerce at The US-China Business Council; http://www.uschina.org/statistics/tradetable.html

23 Horst Köhler, Managing Director of the International Monetary

firms from technologically advanced countries. A small 'infant' firm in an industry on the domestic market would have little chance of successfully competing against the larger international firms which have enormous advantages in size, market share, production knowledge etc. Meeting the international firms (which have far lower average costs) in head-on competition is hardly doable, whereby some form of intermediate (= transitional, temporary) protection can be warranted. This is the **infant industry argument** for trade barriers, which prescribes reducing price and or cost differences between small domestic firms and international firms. Commonly a mix of tariffs, import quotas and/or subsidies are used to protect the infant industry, allowing these firms to move along the learning curve and ultimately be able to employ comparative advantages and compete with older and more experienced foreign producers.

The argument is frequently – if not always today – put forward in the context of developing countries competing on international markets ruled by firms from developed countries. When a less developed country (LDC) initially cannot utilise its possible comparative advantage in, say, manufacturing then tariffs can be used to raise domestic market prices and allow domestic firms to cover their costs – which are higher than international producers'. A subsidy to domestic producers would have the effect of lowering domestic production costs. In both cases the domestic firms become more competitive and increase their output, enabling the industry time to 'grow up' under the cover of temporary protectionism. When the LDC becomes increasingly efficient and competitive, the tariffs/subsidies are gradually reduced to free trade levels.

Pro and con in infant industry argument

Pro: It should be pointed out immediately that the three major industrialised nations, US, Japan and Germany, all had high trade barriers in their initial stages of industrialisation. There is a case to be made in that the *dynamic gains* to the domestic economy – the 'spin-off effects' of moving along the learning curve such as increased know how, technological advancement etc – ultimately provides benefits which outweigh such intermediate costs as higher prices and lower consumption. There has also been some justification for adopting infant industry protections using market failure arguments; the domestic capital markets in many LDCs have simply lacked viable financial institutions and funds to finance investment in industries with potential domestic comparative advantages. Thus there is an *economic argument* in support of infant industry protectionism if the

social benefits (comparative advantage, increased knowledge/technology, improved workforce) are greater than the losses (deadweight losses due to higher prices and misallocation).

Con: There are, however, few absolute successes in using infant industry protection according to most economic analyses. The criteria for 'absolute successes' would be that the protectionist measures were indeed *temporary* and have *benefits which exceed the costs* to society. Historically, few tariffs and subsidies implemented in line with the infant industry argument became anything like 'temporary' but in fact became permanent for longer periods – often at the behest of (large and often monopolistic) domestic suppliers fearing competition. High tariffs and/or domestic subsidies all too often create comfort zones where even inefficient domestic firms can stay in business. As for costs and benefits, governments have been noticeably poor in picking winning industries to protect and support. For example India was long one of the most protectionist countries in the world, and had almost impenetrable barriers on imported cars. Yet when the country introduced sweeping changes to foreign trade policy in 1991, following in the footsteps of China ten years earlier, India became a major exporter in entirely different areas, noticeably textiles, leather goods and more recently Internet-based services such as call-back centres.

Anti-dumping barriers

> "If the other fellow sells cheaper than you, it's called dumping. 'Course, if you sell cheaper than him, that's mass production." Will Rogers

When domestic suppliers sell goods abroad at below domestic production costs (HL: below *marginal* cost – assuming a competitive market) or domestic market equilibrium price one says that **dumping** takes place. While it is possible that firms might do so in order to get rid of excess goods, use excess production capacity or capture market share by driving out rivals (so called 'predatory pricing'), dumping far more commonly arises due to *export subsidies*.

Definition: 'Dumping and anti-dumping tariffs'

When a country's firms sell goods below production cost (HL; MC) – and more widely defined; below the domestic price level – the goods are being dumped on the importing country. When the importing country taxes these imported goods to reflect the 'true' cost of production one speaks of an anti-dumping tariff. (Note: HL should use '…below the marginal cost…').

Problems arising in setting anti-dumping tariffs

The WTO allows for anti-dumping tariffs to offset an unfair price differential and thus trade advantage of foreign firms. Two immediate problems arise in setting anti-dumping tariffs:

1. It is most difficult – in some cases impossible – to correctly **assess the production costs** of firms accused of dumping. Often these difficulties arise from the fact that there are subsidies given to firms in the dumping country which are immensely difficult to accurately account for.

2. As a result of the above, it is common to use the domestic market price of the country of origin to determine whether dumping is taking place. This creates a problem for firms which are simply **adapting to local market prices** in furtherance of gaining market share.

A good number of countries consider dumping as an unfair trade practice and so too do WTO rules; offending imports can be taxed to bring the import price up to what is considered the correct market price in the importing country. Therefore anti-dumping tariffs are allowed by the WTO, resulting in accusations of unfair trade barriers by the countries accused of dumping. And since all anti-dumping tariffs must be instigated by an official complaint to the domestic regulatory body by firms in an industry, the counter-argument of the country being dumped upon is that domestic industries are simply being protected from unfair competition. Many of the disagreements herein are due to the difficulties in assessing whether dumping is in fact taking place, since it is so difficult to evaluate the actual costs of production in the exporting country. This creates one of the major obstacles to free trade; since real dumping is so difficult to pinpoint and prove one way or another, countries can slap anti-dumping measures on goods much like any other tariff. Anti-dumping tariffs are also easy to defend politically using the argument of 'fairness' in trade. [27]

One very high-profile round of dispute in international trade concerned the US imposition in 2001 of up to 30% anti-dumping tariffs on steel which the EU violently protested and which the WTO concomitantly found to be illegal in 2004,

allowing the EU to impose retaliatory tariffs of $2.2 billion on US goods from March 2004 onwards. In December 2003, President Bush II averted what started to look alarmingly like an all-out trade war by rescinding the tariffs. [28]

During the onset of the world financial/liquidity crisis in 2007, the WTO organisation reported a surge in anti-dumping investigations. During the first half of 2008, there was a 39% increase in initiations (e.g. anti-dumping complaints) by WTO members for the WTO to investigate alleged dumping. Turkey, USA, India and the EU were the top initiators. [29]

You will have noticed the strong bias in my use of examples, where anti-dumping policies involving the US and EU make up the lion's share. This is in no way accidental as 90% of all protectionist measures initiated world wide between 1998 and 2008 with the WTO were in fact anti-dumping measures and the US and EU together accounted for well over half them. [30] As for the success of anti-dumping regulations imposed, 71% of anti-dumping claims in the EU and 80% of those in the US succeeded between 1980 and 1997. I must say that I was most surprised at how anti-dumping policies in the US *overwhelmingly* surpassed all other countries in the world and a few minutes of investigation provided the answer; two US laws – one from 1916 and one from 2000 – grant US firms not only the right to lodge complaints about dumping but also **grants these American firms the proceeds of fines** levied on firms found guilty of dumping. In other words, there is a massive incentive for domestic US firms to file grievances of being dumped upon by foreign firms. The WTO has ruled that the US legislation is in violation of trade rules but there has as yet been no compliance by the US.

Strategic arguments

Referring to any number of conflicts and wars, I often claim that the only thing that goes up faster than anti-aircraft turrets in a country entering a war are trade barriers – both in the form of coastal mines and forced embargoes put in place by enemies and their allies. A country which has specialised greatly will have difficulties in getting hold of traded goods during times of conflict, resulting in any number of goods – for example food and clothing – having to be produced domestically in order for the war-time economy to function.

---

27 Anti-dumping duties have in fact been referred to as 'ordinary protection with a great public relations programme'. See McCulloch, *Trade liberalization and poverty*, page 295

28 BBC News at http://news.bbc.co.uk/1/hi/business/3291537.stm

29 WTO: 2008 PRESS RELEASES. Press/542. 20 October 2008, *Secretariat reports surge in new anti-dumping investigations –* see www.wto.org

30 http://www.ecipe.org/media/publication_pdfs/ten-years-of-anti-dumping-in-the-eu-economic-and-political-targeting.pdf

TRADE

This non-economic argument has a degree of general validity, where pure military goods will naturally have the strongest argument in favour of domestic production and protectionism since no country will want to be dependent on potential enemies for the supply of vital military goods. It is this 'self-sufficiency' argument which has led a number of smaller countries to become exporters of military goods; in order to gain benefits of scale, military suppliers will sell to foreign countries – and not uncommonly receive domestic export subsidies in direct support. This lowers the average cost of goods produced for the domestic market and the export revenue can even help cover some of these costs.

Source of government revenue

Two centuries ago, the largest share of government revenue came from tariffs. These customs duties were easy to control since points of access (i.e. ports) were limited, and also easy to collect as the goods had to pass customs warehouses before being distributed. Today, however, modern administrative systems, well functioning census bureaus and efficient tax offices allow other forms of taxes, such as income taxes, to form the brunt of government tax revenue.

However...

Noticeably different in this respect are less developed countries, LDCs, where domestic production in many cases will not provide viable substitutes for imported goods – primarily manufactured goods (secondary goods) with higher value-added content. The contrast in levying tariffs on such imports is shown in diagrams I and II (Figure 65.5) which show how the government in a more developed country (MDC) will have less to gain from tariffs than an LDC.
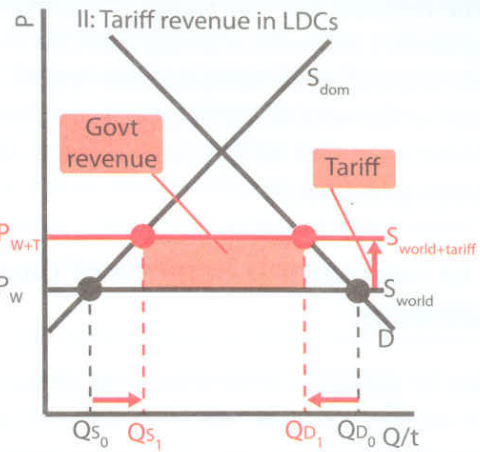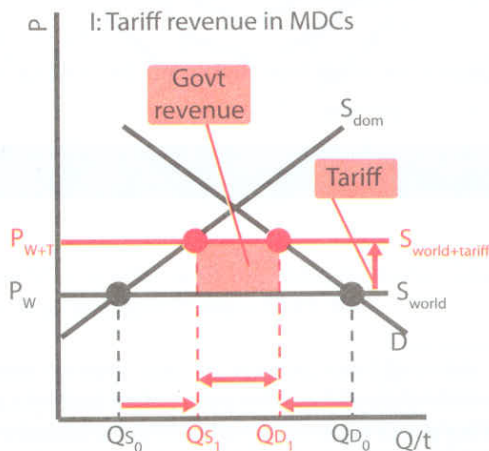
Figure 65.5 Tariff revenues in MDCs and LDCs

An MDC, Figure 65.5, diagram I, will have a high level of industrialisation, intra-industry trade and therefore viable domestic substitutes, which means that both domestic supply and demand will be relatively elastic.

- The situation in an LDC, diagram II, is often the opposite; a low proportion of intra-industry trade, relatively low level of industrialisation compared to MDCs and negligible domestic substitutes with resulting demand inelasticity.

The imposition of a tariff will have a greater impact on tax receipts in an LDC than an MDC – shown by the size of the shaded area in the two diagrams. The above helps to explain why average tariff rates in developed countries are less than 5% while developing countries have an average tariff of roughly 20%.[31]

Additionally, the formal economy in LDCs is frequently relatively small and the informal economy does not comprise a taxable base. Tariff revenues in developing countries therefore make up a far larger portion of tax receipts and general government revenues. The United Nations Conference on Trade and Development (UNCTAD) estimates that close to 18% of all government revenues and over 20% of total tax receipts come from import duties in developing countries. Corresponding figures for developed countries are less than 4% of both

31 Hertel, Thomas W & Martin, Will, *Would developing countries gain from inclusion of manufactures in the WTO negotiations?* Geneva (WTO) 1999, (WTO/World Bank Conference on Developing Countries' in a Millennium Round, 20–21 september 1999), see http://www.itd.org/wb/hertel.doc and Environmental Working Group (EWG); an independent US organisation monitoring subsidies, at http://www.ewg.org

government revenues and tax receipts.[32] Understandably, a developing country government lacking a distinct domestic tax base, a well functioning tax/ administrative system, and thus a steady inflow of tax receipts will be quite dependent on tariffs. Unfortunately, some 70% of this tariff revenue comes from other developing countries.[33]

Means to overcome a balance of payments disequilibrium

The **balance of payments** is an account of all the monetary dealings a country has with other countries and basically accounts for all money flows across the border during a period of time. The balance of payments is comprised of two parts; the current account and the capital account. The **current account** shows export revenue and import expenditure while the **capital account** shows money in- and out-flows resulting from foreign investment and loans. Leaving further details for Chapter 70, a balance of payments disequilibrium refers to a situation where export revenue for a country is less than import expenditure. This is known as a **trade deficit**, shown below in Figure 65.6, where export revenue is $400 million and import expenditure is $500 million. Disregarding other money flows (for example arising from services and transports) there is a current account deficit, i.e. a balance of payments disequilibrium.

Current account

X revenue: $400 million

M expenditure: $500 million

Trade balance: $-100 million

Current account deficit: $-100 million

Disregarding other in- and outflows for the time being, the above trade balance results in a current account deficit.

Figure 65.6 Simplified current account

If government imposes tariffs (known as an *expenditure switching* policy) an amount of domestic expenditure will be diverted away from imports toward domestic goods. The decrease in total spending on imports improves the trade

balance and thus improves the current account in the balance of payments.

However...

It is also possible, perhaps even likely, that countries on the receiving end of protectionist policies *retaliate* in kind – in much the same way that the Great Depression of the 1930s was exacerbated. Such 'reciprocal' or 'retaliatory' trade barriers are often the result of *beggar-my-neighbour* policies[34]. It is therefore a distinct possibility that the improvement in the current account will be eroded in the longer term. The frequency and debilitating effects of such tit-for-tat protectionism led the WTO in 1995 to outlaw the practice of balance of payments readjustment using tariffs. Yet, as Chapter 69 will show, it is still possible to use exchange rate adjustments to similar effect.

POP QUIZ 65.1

1. Compare the costs and benefits of using tariffs and subsidies as a method of protecting domestic industries.

2. Why might the 'infant industry argument' be considered a viable trade argument?

3. Criticise the argument that trade barriers can save domestic jobs.

4. Explain how a non-economic barrier to trade can limit imports.

5. When might protectionism be labelled a 'beggar-my-neighbour' policy?

6. Explain how trade barriers might be used to adjust the balance of payments.

---

32 *Tariffs, Taxes and Electronic Commerce: Revenue Implications for Developing Countries,* Susanne Teltshcer at UNCTAD, 2000.
33 Hertel, Thomas W & Martin, ibid.

34 Beggar-my-neighbour policies are any type of protectionist foreign trade policy where domestic output and jobs are 'saved' by limiting goods exported by others – i.e. at another country's expense. Tariffs, quotas and currency devaluations are common examples.

TRADE

The case against protectionism

The remainder of this chapter is partially a reiteration of the counter-arguments put forth above using real world examples to back up economic theory. Most of the arguments against protectionism centre on efficiency losses and higher prices, i.e. the allocative and welfare losses to society. Perhaps the most obvious example of damages inflicted on society by implementing trade barriers is the Great Depression in the 1930s. When US President Hoover signed the **Smoot-Hawley tariff bill** in 1930 in order to protect American jobs, average American tariffs increased to 53%. Most economists agree that this worsened a world economy already in serious trouble. The US was even then the world's largest importer and the bill severely hampered income in exporting nations – many of which were in desperate need of US dollars to pay off war debts. The bill sparked massive retaliatory tariffs from over 25 countries, setting off a tit-for-tat escalation of trade barriers. By 1933 world trade had fallen by approximately 25% and industrial output by 32%. The bill had no positive effect on American unemployment; 3 million people in the US were jobless in 1930 and 13 million by 1933.[35] Figure 65.7 shows how world trade shrank over the five year period between 1929 and 1933.

World Trade 1929-33, $bn

—1929 —1930 —1931 —1932 —1933
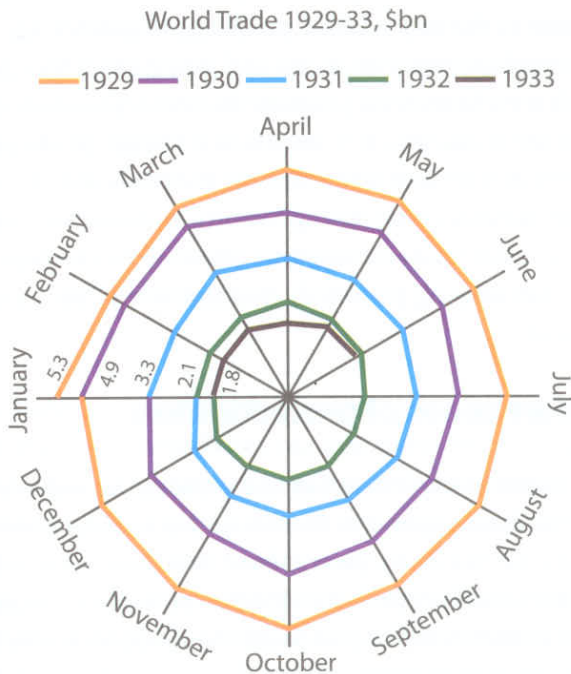
Figure 65.7 Spider web of protectionism   Source: League of Nations Survey 1932-33

35 *A world without walls, Mike Moore*, page 27; National Center for Policy Analysis at http://www.ncpa.org/oped/bartlett/oct2999. html; and Canadian Dept. of Foreign Affairs and International Trade at http://www.dfait-maeci.gc.ca/ciw-cdm/wto1-en.asp

During February 2009, there were some very frightening events unfolding in the international trade arena with the US at the forefront. The on-going financial/liquidity crisis is causing old-fashioned job-saving protectionism of the 1930s' type to rear its ugly head. A G20[36] meeting in November 2008 resulted in promises not to raise existing or implement any new trade barriers – within a week both India and Russia raised tariffs on, respectively, cars and steel.[37]

In the US, the current trend is falling trade, increased import restrictions, and very distasteful elements of mis-guided patriotism: 'Buy American' addendums are being included in economic stimulus bills. One such bill passing through the US House of Representatives in January 2009 was a USD825 billion stimulus package which included a requirement that only US steel be used in any public works projects. A study by the Peterson Institute for International Economics estimates that this would create about 1,000 new jobs but cost the US far more in lost jobs due to retaliatory measures by other countries.[38] The European Union (EU) very quickly warned the US that the Buy American provisions would be in possible violation of WTO rules.

Inefficiency in resource allocation

A few years ago on my Greek island, Sifnos, I hitched a ride with a very pleasant Greek gentleman in his 40s who was driving his jeep off to a rather remote spot. I tagged along and after 20 minutes of rough-riding on what essentially were donkey paths, we arrived at a cul-de-sac on a slope overlooking the sea. The Greek gentleman got out and started putting on rubberized overalls (it was 40°C in the shade!) and a backpack that looked like, well, a vacuum cleaner with a nasty spray nozzle attached – something right out of *Ghostbusters*. He was on his way down to spray his newly planted olive trees and since he lived in Athens he came out only once a year and gave them a really good spraying – thus the heavy protective gear. It turned out that he had a number of plots on various islands where he had a few hundred olive trees.

Standing discretely upwind of him and his toxic backpack, I asked if it really made economic sense to have bits of land all over the Aegean Sea – especially for olives for which there was

36 The G20 is comprised of the eight largest industrialised countries plus eleven smaller industrialised countries and emerging market nations. See http://www.g20.org/about_index.aspx

37 The Economist, *Fare well, free trade*, Dec 18th 2008

38 Dan DiMicco, chief executive of Nucor (a large US steel manufacturer) dismissed the report as "complete garbage". (See for example International Herald Tribune, February 4 2009 *'Steel-state lawmakers vow to save 'Buy American'*

TRADE

such a mountain of excess already. He laughed uproariously and then explained how olives were subsidised by the *tree*, not *per kg* of actual olives produced. It turned out that he had taken bank loans and bought up a bunch of out-of-the-way plots on various islands in order to plant as many olive trees before impending changes in subsidy rules went into effect. The subsidies he collected for each tree more than made up for the costs of servicing his loans and visiting various land plots for a yearly spraying of insecticide. He looked to pay off his loans in about 20 years, by which time he would be the owner of quite a few hectares of land. Not once did he mention actual olive or olive oil production.

Story-time!
First Class Cows

The total cost of agricultural subsidies in the 29 OECD countries in 1999 was $US361 billion. This was enough to send all 56 million OECD cows 1.2 times around the world on first class flights. If the cows were willing to accept business class, each bovine traveller would have $US2,800 in pocket money left over to spend at the taxfree shops along the way[1]. And we worry about cows being mad.

1. Taken from Till världskapitalismens försvar, Norberg , page 141

"What? You need a cigar for the first class lounge at the airport?"

The above story is a somewhat extreme illustration of the often horrendous wastefulness and downright misuse of resources resulting from trade barriers – in this case subsidies. Recall that optimal resource allocation is using limited factors of production to create the best possible outcome for society – 'best' being defined as the point rendering the highest output at the lowest cost. No waste is present, i.e. no goods would be made without them being in demand. As such, when supply equals demand, the market has cleared and there is optimal allocative efficiency. At any other level of output, resource allocation will be sub-optimal.

In addition to being directly wasteful, tariffs and quotas have other economic costs attached, such as administration, enforcement and regulation costs. Thousands of civil servants are needed to process applications and send out payments. For example, in preparation for joining the EU in May 2004, the Polish bureaucracy expressed concern in having to hire an additional 4,000 workers simply to administrate the sending of payments to eligible farmers. As for the costs of enforcement, one could mention the costs of sending police officers out to olive groves to count olive trees as I've seen done in Spain.

Subsidies to domestic industries not only cost taxpayers billions every year but seriously distort production and reallocate resources to far less efficient producers. Subsidies are often used in conjunction with quotas and minimum price schemes – which serve to increase prices for domestic consumers rather than lower them. When more efficient foreign producers are disadvantaged by such trade barriers, the protectionist policies will have the effect of increasing domestic costs and/or domestic prices. In either case, the domestic economy moves away from the market clearing price with resultant suboptimal resource allocation.

Comfort zones of protectionism

When countries implement barriers to trade in accordance with any of the above arguments, there is seldom an 'Expiration date' attached. All too often, protectionist measures become long-lived and virtually permanent. Apart from the direct monetary costs to governments/taxpayers and the allocative losses, there are a number of **dynamic efficiency costs** which arise behind protectionist walls. Many of these costs are similar to those that might arise in monopolies – and indeed trade barriers create monopoly outcomes in domestic markets on a macro scale. The long run costs of industries dependent on protectionism are summarised in three points:

- **Misallocation costs:** Trade barriers cannot in the long run save non-competitive (sunset) industries from competition. By imposing tariffs on imports and/or subsidising domestic firms, the government is simply delaying a process of reallocation at great cost to society in terms of alternative output. Protectionism creates overproduction – often at higher prices to consumers – and inhibits inefficient firms from reallocating resources to other sectors.

- **Disincentive costs:** When domestic firms come to rely on protectionism, the propensity to improve productivity is diminished. Ultimately other – non protected countries – will increase productivity and lower production costs to the point where they will be competitive even facing a tariff. Additionally, protected industries will have a lower level of research and development (R&D) spending and technological innovation due to lack of international competition. The secondary effects are that industries in non-protected industries will ultimately have higher costs compared to foreign firms, losing competitive edge and furthering the productivity gap between domestic and foreign firms.

- **Forward linkage effects:** When imported raw materials and components become more costly for firms using these inputs, the production cost moves forward down the line towards end-users. For example, when a tariff is put on steel, there is an increase in the price of steel and therefore an increase in the cost of producing trucks → costs for transportation firms rise → cost of shipping Widgets rises → price of Widgets rises.

The forward linkage effects of protectionism can in fact *cost* jobs rather than *save* them. When manufacturers of cars, washing machines and refrigerators lose customers – both domestically and abroad – there will be layoffs. Several studies in the US have shown how tariffs on imported steel would save steel jobs *at the expense* of jobs in steel consuming industries. Steel consumers outnumber steel producers by 40 to 1 and each steel job saved would cost between 8 and 10 jobs elsewhere down the line of manufacturing.[39]

Increased prices of goods and services

No matter how one slices the cake, one piece will fall on the floor; every form of protectionism has a price tag which is ultimately borne by consumers/taxpayers/citizens. Tariffs and quotas cost consumers via increased domestic prices directly, and subsidies indirectly via taxes. Philippe Legrain, an economist and free trade advocate, estimates that American citizens have paid about $US174 billion in total for steel protection between 1969 and 1999; $US151 in higher steel prices plus an additional $US23 billion in taxes for steel subsidies.[40]

Another extreme example is in textiles and clothing, where the cost of protectionism (tariffs, quotas and subsidies) to American consumers is estimated at $US24 billion annually, preventing the loss of 170,000 jobs - a cost to US taxpayers of $US140,000 per textile job.[41]

While the most obvious examples of how protectionism raises consumer prices on both domestic and imported goods are to be found in tariffs and quotas many other forms of protectionism render the same outcome. When domestic suppliers can rely on various forms of administrative, technical and legislative barriers to imports, there will be a strong incentive for domestic firms to increase their profit margins by heavy mark-ups (= price increases) on retail sales. My personality being such that I am prone to opening eggs with a sledgehammer, I will exemplify using possibly the most extreme example; Japan. The domestic price in Japan (over and above world market price) for rice is 734%. Allowing for a reasonable profit for an importer, the average Japanese family is thus spending close to 7 times more on this staple food than is necessary. Similarly, the Japanese above-world-market price on wheat is 477% and on radios and televisions 607%. Guess what the tariff level on these goods is. Zero percent. For a number of other goods, the above-world-market price is considerably higher than the tariff, for example, tea and coffee prices are 718% higher yet tariff levels are 12%. The above-world-market-price on cosmetics is 661% (2% tariff) and on gasoline 229% (5.5% tariff). The estimated total cost of protection to Japanese consumers is an estimated $US110 billion a year – 3.8% of national income.[42]

Effect on domestic competitiveness

It is quite possible that domestic production suffers from increasing costs and thus decreased competitiveness due to protectionism. There are two ways this can happen.

---

39 See *Romancing big steel*, The Economist February 14th 2002; and *A world without walls*, Mike Moore, page 30

40 *Open World...*, Legrain, page 34
41 Irwin, page 93
42 Perloff, page 300

Firstly, there is the **forward linkage effect** given above, where higher domestic production costs will hamper the domestic economy's ability to compete on international markets. An economics adage from the US in the 1980s read; "A tax on Nippon Steel is a tax on General Motors!" What happened was that the steel industry in America lobbied for tariffs on cheap imported steel from Japan. When such tariffs were in place, domestic American suppliers were able to increase their prices of steel – which increased the cost of producing automobiles in America. Guess what consumers around the world did? Yes; they increased their purchases of Japanese *cars*. The forward linkage effect of higher input prices due to tariffs cost the US economy far more in loss of both domestic and foreign sales of cars than was gained by slapping tariffs on imported steel. In extreme cases, the production cost disadvantage will lead firms to *relocate* production outside protectionist walls, i.e. to another country. For example, high US tariffs on sugar – a key ingredient – caused one of America's largest food producers, Kraft, to move production of the classic LifeSavers candy from the US to Canada.[43]

Secondly, the protectionist policies **shield domestic producers from competition** which lowers there ability to compete with more open economies which are forced to forever enhance productivity, improve products and innovate. Domestic suppliers allowed to lead 'the quiet life' sheltered behind protectionist walls will have a marked tendency to increase prices domestically and this abnormal profit will ultimately build in inefficiency. Lack of competitive forces will also create a degree of R&D lethargy and unwillingness to invest in new technology. Taken together, domestic firms will see a continued erosion of cost advantages towards foreign firms under the false comfort provided by trade barriers, as competing firms in free-trade economies continuously strive to become more efficient, cut costs and gain scale benefits.

Corruption

Go back and re-read the parts on Indonesian barriers to trade. It will surely come as no surprise that the more rules, regulations, limits and indecipherable paragraphs a country has, the more room there is form the official on the other side of the desk to hold you for a little palm-grease, i.e. a bribe. It is so common that when governments in countries I have lived implement any type of new trade barrier or restriction, most of the expatriates say something like "Well, the government is giving the customs officials another raise!" In fact, it is rather cunning; new rules provide the officials with more room to extract bribes and thus increase their incomes – without this causing a burden on

---

the government budget. One can view it as a form of tax on consumers really.
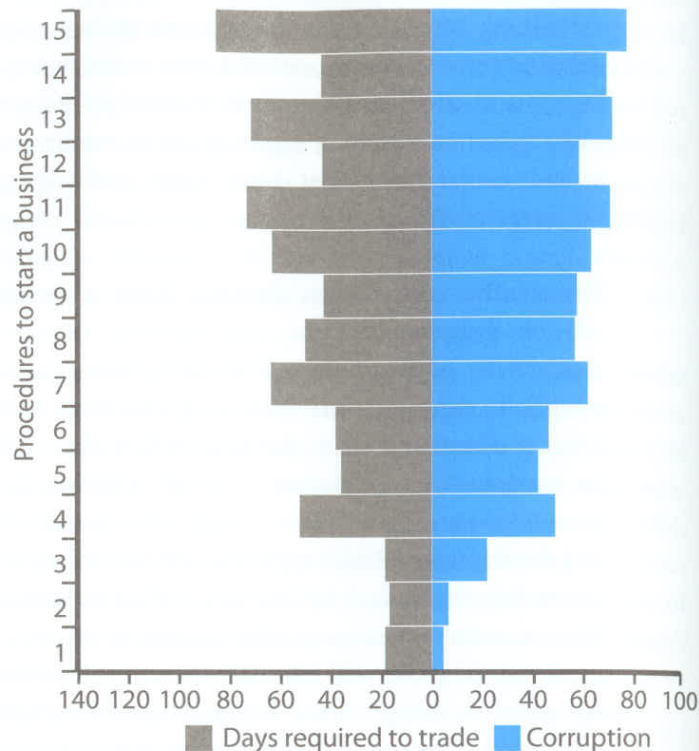
Figure 65.8 Correlation between barriers to trade and corruption

Source: World Bank, Doing Business and Worldwide Governance Indicators for 2008

A study by the World Bank in 2008 indeed indicates positive correlation between the difficulties in international trade and the level of corruption (figure 65.8). Note again that correlation does not indicate which way the causal flow is moving! It is quite probable that protectionism causes corruption…and also that increased corruption causes increasing barriers to trade. The reason for this is that when officials take bribes instead of simply collecting correct import duties, governments lose money. This in turn can force governments to take even further protectionist steps due to large black market activity and loss of tariff revenue.

---

43 *A world without walls*, Mike Moore, page 47.

TRADE

OUTSIDE THE BOX

Final rumination on trade barriers

In light of the considerable economic and welfare costs involved in trade barriers, it seems to be something of a lunacy to have so many forms of protectionism in place. Few of the arguments upholding protectionist policies are regarded by economists as viable in terms of economic welfare, so why are countries given to imposing tariffs, quotas and many other forms of barriers? Even better, why are so many obviously senseless policies implemented by very sensible policy makers? Quite frankly, I really don't know – and it's something to which I've given considerable thought and study.

In a dark moment I think: Domestic firms, labour groups and other groups have a vested interest in protecting themselves from foreign competition. These groups are well organised and have strong well-funded lobby groups which have long since been able to sway politicians in their favour in spite of the fact that these groups are in a minority. Populism, vote-winning and simply staying in power are powerful incentives for politicians to listen to anti-trade slogans, which ultimately has resulted in an 'anti-trade litany' which few dare to dispute – at least in public. Furthermore, history has shown how easy it is to gain political power by pandering to the unfortunate undertows of xenophobia (= fear of foreigners/ things foreign) and nationalism.

I also wonder if the media is not propagating the myth of 'harmful trade' by exploiting peoples' fear and ignorance. It takes 10 seconds to put forward the argument that jobs are being lost to foreign countries and that domestic citizens will be worse off. This provides a wonderful 20 second sound-bite on TV and a splashy newspaper-selling headline that 12 year olds can understand. Explaining the theory of comparative advantage, on the other hand, takes time and some real intellectual effort to grasp and accept – and doesn't make much of a story either in terms of sensationalism. So when television stations set programming, the 30 minutes of evening news will give 45 seconds to the story '1,500 jobs lost to Poland/China/ Vietnam' while 'How everyone ultimately gains from trade' will not get the minimum two hours of TV network time needed to explain it.

It also increasingly strikes me that my money is a vote and if I vote for imports, well, that's my business. Any democratic and liberal society has strict rules on the ballot box; one's vote is secret to avoid getting blackmailed or otherwise bribed into voting for a particular candidate or party. Tariffs and other forms of protectionism are basically *anti-democratic* since they are attempting to force me to vote for goods I don't want.

In a bright moment I think: Most of us ordinary people quietly vote for trade every time we add to our marginal utility in selecting from a wide variety of imported goods and for us the benefits are quite evident. There's a wonderful Swedish adage; *Hälsan tiger still*, which is a bit tricky to translate but in essence reads 'Health lies quietly still'. The millions of consumers around the world who see immense increases in their standard of living and quality of life due to trade will have little need to manifest this in media proclamations or 'Thankful Consumers for Trade' organisations. They have stated their case with their wallets.

I then posit that most of the anti-trade and anti-globalisation rhetoric has its origin in various interest groups who stand to be disadvantaged by trade, such as people losing jobs when firms move abroad and firms that lose revenue and market share to foreign competitors. I wonder if a great deal of the vocal anti-trade protests in fact aren't simply due to well organised and well funded interest groups voicing opinions which a vast majority of people wouldn't actually sympathise with given the opportunity to seriously weigh the pros and cons of trade. Perhaps protectionist tendencies are allowed to get a strong foothold in trade policy simply because a silent majority is not voicing an opinion.

In my brightest moment I think: Thank God for my students and colleagues around the world! What better way to set a few shocking wrongs to right than by a sound education in economics. Please, study hard; get excellent grades; go to university and then get into a position of power where you can influence society for the better. Oh, and pay lots of taxes so that my colleagues and I get ample public/merit goods and pensions. I figure we deserve it.

TRADE

Preparing for exams

Short answer questions (10 marks each)

1. How might a country improve its comparative advantage by producing manufactured goods?

2. Explain how a country might use trade barriers in order to improve the current account in the balance of payments.

3. "Trade is much freer today since tariffs have fallen drastically in the past 50 years." Discuss this statement.

4. Explain why subsidies to domestic firms act as a trade barrier.

5. How can trade barriers for agricultural goods in developed countries affect developing countries?

6. What factors might influence the competitiveness of a firm operating in export markets?

7. Use a diagram to illustrate how a tariff causes resource misallocation.

Extended response questions (25 marks each)

1. a. What is meant by 'dumping'? (10 marks).

   b. Why have anti-dumping tariffs become so common and how might manufacturers in low cost countries defend themselves against dumping accusations? (15 marks).

2. a. What are the possible advantages of free trade? (10 marks).

   b. In light of these advantages, why do so many barriers to trade exist? (15 marks).

3. a. Identify the ways in which an economy can limit imports. (10 marks).

   b. Argue for and against such policies. (15 marks).

4. a. Distinguish between the terms of trade and the balance of trade for a country. (10 marks).

   b. How might a country use trade barriers and the exchange rate to affect both the terms of trade and the balance of trade. (15 marks)

Summary & revision

1. Protectionism is any deliberate action by a government which lowers the ratio of export prices to import prices.

2. A **tariff** is a *tax on imported goods* and has the effect of raising the price of the good, lowering the quantity of imports and increasing the domestic quantity supplied.

   a. Consumer surplus decreases

   b. Domestic supplier surplus increases

   c. Government receives tariff revenues

   d. Foreign suppliers' revenue falls

   e. There is a deadweight loss showing suboptimal resource allocation

3. A **quota** is a *physical limit on imports*. It has the same effect as a tariff but instead of tariff revenue there is a windfall gain (unexpected increase in revenue) for foreign suppliers.

4. A **subsidy** is a *gift or money grant to domestic suppliers* which increases domestic supply. There is no loss of consumer surplus but a cost for government and an efficiency loss.

5. **Administrative trade barriers** can consist of bureaucracy, red tape and extensive regulatory demands on imported goods. They slow down the market entry process and increase costs for foreign suppliers.

6. **Health and safety** act as trade barriers as they often consist of rules and regulations which limit goods market entry and raise import costs. They have been frequently misused.

7.   **Environmental standards** – such as emission limits on cars and requirements for certified plantation grown hardwoods – limit some trade and are mostly viable arguments for trade barriers.

8.   The most common **arguments for protectionism** are:

a.   *Protection of industries and unemployment* (historically not a viable argument and frequently mis-used politically)

b.   *Infant industry argument* which means that domestic industries are protected (via tariffs often) against foreign multinational companies which have large economies of scale. The protection is to be removed when the domestic industry is strong enough to stand on its own feet.

c.   *Anti-dumping tariffs* are justified by the claim that foreign producers are selling at below production costs (HL: MC) and that a tariff 'evens-out' the playing field. This form of protectionism is one of the most commonly used today and is widely thought to be misused.

d.   *Strategic arguments* claim that a country needs to have a reasonable degree of self-sufficiency in case of conflict, war or natural disaster. This argument is now largely obsolete.

e.   Tariffs comprise a source of government *tax revenue* and constitute the brunt of tax revenues in many LDCs.

f.   Trade barriers can be used to adjust a *balance of payments disequilibrium*. The danger is that reciprocal barriers set up by trade partners might in fact ultimately lower tax revenues due to falling exports and GDP.

9.   The main **arguments against protectionism** are:

a.   *Inefficiency* (see deadweight loss triangles in the diagrams) and *misallocation* of resources

b.   Possibility of *reciprocal barriers* which ultimately can cost more jobs tan it saves.

c.   Protectionism creates 'comfort zones' for domestic producers and this reliance causes dynamic

efficiency *losses* such as disincentives to improve *productivity*, lower levels of *innovation*, *poor quality* goods and the *forward linkage effects* on domestic producers in having to pay more for imported factors of production.

d.   Domestic consumers suffer from *higher prices*, *limited choice* and *lower quality goods*.

e.   *Protectionism creates corruption* when complex rules give room for bureaucrats and civil servants to extract bribes, such as allowing things through customs and speeding up paperwork.

TRADE

465

66. Calculations of Effects of Tariffs, Quotas and Subsidies

Key Concepts: HL extensions

- Calculating changes in prices, consumer and supplier surplus, domestic and foreign producer revenue, government costs/revenue, efficiency loss, net surplus loss, deadweight loss
  - Tariff
  - Quota
  - Subsidy

H L students will have to put numerical values on all the areas in the core trade barrier diagrams. It might be a good idea to quickly revise Chapters 4 to 8.

Calculations

The diagrams in Figure 66.1 represent the effects of a tariff, quota and subsidy. All the diagrams are indexed in price/quantity at original imports and the areas resulting from protectionism are lettered. I have altered the PED and PES in the diagrams so as to enable you to differentiate between the numerical values....and of course to force you to do additional calculations. Calculate the areas and check your answers with the cheat-sheet given in the end of chapter revision.

II: Quota calculation
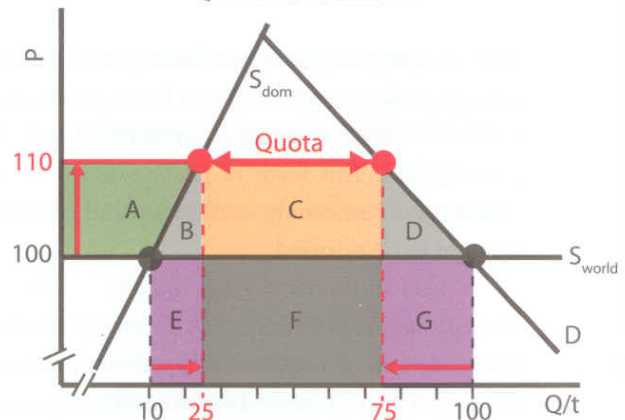
III: Subsidy calculation
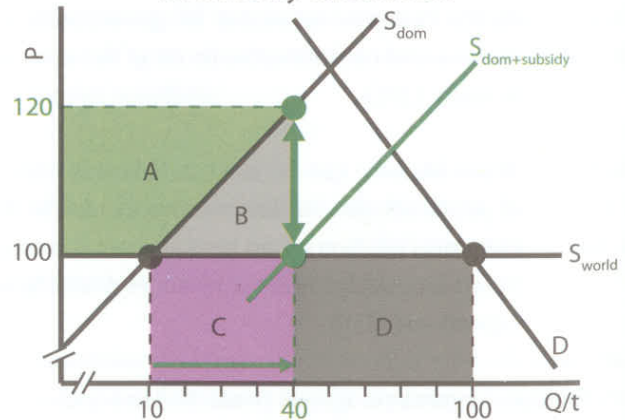
Figure 66.1
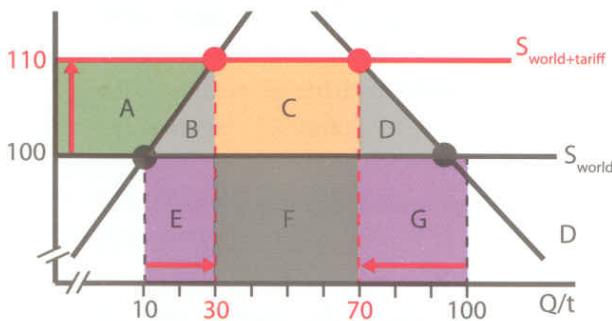
Tariff

The tariff is 10% in Figure 66.1 (diagram I). Calculate both the area and the numerical values for:

1. Gross loss of consumer surplus

2. Increase in domestic suppliers' revenue

3. Loss of foreign suppliers' revenue

4. Remaining foreign suppliers' revenue

5. Efficiency loss due to tariff (e.g. units produced post-tariff where domestic suppliers' MC > MC_{world})

6. Net loss of consumer surplus

7. Dead weight loss

8. Government tax revenue

Quota

The quota in Figure 66.1 (diagram II) has similar effects – with one exception worthy of mentioning, namely that there is no government revenue. Instead there is a windfall gain to foreign suppliers, e.g. an unexpected increase in revenue. (Note: this is not necessarily a net increase in foreign suppliers' revenue! It depends on how large areas E and G are.) Calculate

1. The size of the quota (e.g. units rather than money values)

2. Increase in domestic suppliers' revenue

3. Loss of foreign suppliers' revenue

4. Remaining foreign suppliers' revenue

5. Efficiency loss due to tariff (e.g. units produced post-tariff where domestic suppliers' MC > MC_{world})

6. Net loss of consumer surplus

7. Dead weight loss

8. Windfall gain

Subsidy

The subsidy in Figure 66.1 (Diagram III) shifts the supply-curve and results in a new equilibrium point. Recall the law of incidence '… new equilibrium point → old supply curve → go home …'. Calculate:

1. The subsidy per unit and the total cost of the subsidy to government

2. Increase in domestic suppliers' revenue

3. Loss of foreign suppliers' revenue

4. Remaining foreign suppliers' revenue

5. Efficiency loss due to subsidy (e.g. units produced post-tariff where domestic suppliers' MC > MC_{world})

Before you continue it bears asking; did you actually *do* the calculations for the diagrams above or did you blow through them using a hair-dryer to turn the pages? That is correct; you need to *do* the calculations rather than glance through them. Bruce Lee did not master the front-kick by reading *how* to do it, he did it. A lot.

This is one of those areas where I am frequently asked "Which one is best?" You know the answer; it depends! It depends on how we define 'best' and for whom. Here are a few brief points worthy of pondering…

- In all cases we can see that domestic suppliers have increased their revenues and that domestic supplier surplus increases

- Consumers would probably prefer the subsidy since the price and quantity remains the same – some of the spending simply switched to domestic producers. In tariffs and quotas there is a loss of consumer surplus but not in the subsidy.

- There is mis-allocation taking place in all three examples as illustrated by the deadweight losses due to tariffs and quotas and the efficiency loss due to subsidies.

- Government gains the most in tariffs and the least in subsidies – the tariff brings in revenue and the subsidy costs the government money.

- Foreign suppliers lose…but there is 'loss' and 'loss'. A good with low PED for imports might result in a net

gain (windfall gain) for foreign suppliers when a quota is levied on imported goods.

I often put forward the following three points:

1. *Weak governments use tariffs* since it is the easiest way to collect taxes. No census or complex administration of citizens is needed register income tax. Also, ports, roadblocks at borders and airports are easy to control militarily – giving said weak governments a very inexpensive way to provide the military upon which many of them base their power) ample opportunities for bribes to pad dismally low wages.

2. *Strong governments increasingly use subsidies.* There are some cunning – albeit not very morally appealing – arguments for this. Firstly they can be very sneaky and difficult to prove (such as building new roads next to industrial areas and providing R&D facilities in universities next to technology firms) so this form of protectionism doesn't invite retaliation. Secondly, citizens are fooled into thinking that the subsidies benefit their consumption whilst saving some jobs. Well, true, but not *all* citizens buy the goods being subsidised and the subsidy *does* in fact ultimately have to be paid for by taxpayer monies. Thirdly, the WTO has not been able to implement rules and regulations regarding subsidies – this is part of the dismal failure of the Doha Round of negotiations (not) taking place since 2001. Finally, there is an element of the 'tragedy of the commons' involved, where governments 'counter-subsidise' industries knowing that governments in other countries will do so too. This is largely what is going on in the US and EU within the larger duopoly in passenger airline construction, Boeing and Airbus.

3. My final point is rather cynical but can be reasonably substantiated empirically (ask your TOK teacher) by reading the newspaper on a daily basis. Politicians looking to get into office or re-elected often find that loud proclamations of being 'dumped upon' are real crowd pleasers. Thus there is an incentive for politicians to out-do each other by promising wide-scale protection of jobs in industries and regions by implementing very visible barriers to trade. Basically I am putting forward the opinion that protectionism is often strongly politically slanted rather than based on economic reasoning. I mean, let's face it; the US embargo (= ban on imports) against Cuban goods has been in place since 1960 and has little or nothing to do with economics.

Summary & revision

1. Tariff (see Figure 66.1 (diagram I))

1) Gross loss of consumer surplus: **Areas A to D.**

Calculated as $($10 \times 70) + \dfrac{($10 \times 30)}{2} = 850

2) Increase in domestic suppliers' revenue: **Areas A, B and E. Calculated as TR post-tariff ($110 × 30 = $3,300) minus original TR ($100 × 10 = $1,000) = $2,300.**

3) Loss of foreign suppliers' revenue:

Areas E and G. $100 × 20 + $100 × 30 = $5,000.

4) Remaining foreign suppliers' revenue:

Area F. $100 × 40 = $4,000.

5) Efficiency loss due to tariff:

Area B. Each unit between 10 and 30 has a higher marginal cost (MC) domestically than at S_{world}.

This is an efficiency loss of $\dfrac{(20 \times $10)}{2} = $100.$

6) Net loss of consumer surplus:

Area D. The area is $\dfrac{(30 \times $10)}{2} = $150.$

7) Dead weight loss: **Sum of areas B and D; $250.**

8) Government tax revenue: **Tariff times the imports post-tariff, i.e. $10 × 40 = $400.**

2. Quota (see figure 66.1 (diagram II))

1) The size of the quota (e.g. units rather than money values): **Imports post-quota are 50 units.**

2) Gross loss of consumer surplus:

Areas A to D.

Calculated as $($10 \times 75) + \dfrac{$10 \times 25}{2} = 875

3)   Increase in domestic suppliers' revenue:

   Areas A, B and E. Calculated as TR post-tariff ($2,750) minus original TR ($1,000) = $1,750.

4)   *Gross* loss of foreign suppliers' revenue:

   Areas E and G. $100 × 15 + $100 × 25 = $4,000.

5)   Remaining foreign suppliers' revenue: **Area F and C. $110 × 50 = $5,500. There has been a *net* loss in foreign suppliers' revenue of $3,500 ($9,000 – $5,500).**

6)   Efficiency loss due to quota:

   Area B. Each unit between 10 and 25 has a higher marginal cost (MC) domestically than at S$_{world}$.

   This is an efficiency loss of $\dfrac{15 \times \$10}{2} = \$75$.

7)   Net loss of consumer surplus:

   Area D. The area is $\dfrac{25 \times \$10}{2} = \$125$.

8)   Dead weight loss: **Sum of areas B and D; $200.**

9)   Windfall gain: **Area C. $10 × 50 = $500.**

3.   Subsidy (see figure 66.1 (diagram III)

1)   The subsidy per unit and the total cost of the subsidy to government: **Subsidy per unit is $20 and quantity post-subsidy is 40, areas A and B. Total cost of subsidy to government is $800.**

2)   Increase in domestic suppliers' revenue:

   Areas A, B and C. New TR is $120 × 40 = $4,800 and pre-subsidy TR is $100 × 10 = $1,000. The increase in TR for domestic suppliers due to the subsidy is $3,800.

3)   Loss of foreign suppliers' revenue:

   Area C. $100 × 30 = $3,000.

4)   Remaining foreign suppliers' revenue:

   Area D. $100 × 60 = $6,000.

5)   Efficiency loss due to subsidy:

   Area B. Each unit between 10 and 40 has a higher marginal cost (MC) domestically than at S$_{world}$.

   This is an efficiency loss of $\dfrac{30 \times \$20}{2} = \$300$.

HIGER LEVEL

TRADE

3.2

67. Floating Exchange Rates

Key concepts:

- Definition of exchange rates
- Floating exchange rate regime
  - Appreciation and depreciation
  - Supply and demand
- Determinants of a floating exchange rate
  - Trade flows
  - Relative inflation
  - FDI and portfolio investment
  - Domestic growth
  - Growth in trade partners' economies
  - Speculation
  - Other determinants
  - Deficits and debt
  - Purchasing power parity (PPP) theory
- Effects of the exchange rate on
  - Domestic AD and GDP
  - Inflation
  - Unemployment
  - Balance of payments
  - Current account
  - Government debt

An exchange rate between two currencies is very similar to two goods illustrated by a production possibility frontier (PPF); the cost of one good/currency is always *in terms of the other*. If one American dollar (USD) gets you 0.77 EUROS (€), then **the exchange rate for the US dollar is €0.77**, i.e. the price of the dollar is €0,77. If the US dollar also gets you 0.63 British Pounds (£) then the exchange rate for the dollar is £0.69. Thus: **USD1 = €0.77 = £0.63**.

Definition of exchange rates

In the examples using a PPF, when 1 unit of textiles had an opportunity cost of 2 units of agricultural goods, then the opportunity cost of 1 unit of agricultural goods was 0.5 units of textiles. Similarly, if USD1 = €0.77, then €1 = USD1.3 (1 / 0.77). Thus, using the same figures as above, **the exchange rate for the EURO becomes €1 = USD1.3 = £0.825**. Figure 67.1 below shows the exchange rates for a few major currencies. The exchange rates for each currency are given in columns, i.e. USD1 costs 80.93 Japanese Yen (¥), or 0.92 Swiss Francs (SFR).

Currency April 2012	U.S. Dollar (USD)	Yen (¥)	Euro (€)	U.K. Pound (£)	Swiss Franc (SFR)
U.S. Dollar (USD)	1	0.0124	1.308	1.586	1.088
Yen (¥)	80,93	1	105.84	128.327	88.02
Euro (€)	0.77	0.0094	1	1.213	0.832
U.K. Pound (£)	0.63	0.0078	0.825	1	0.686
Swiss Franc (SFR)	0.92	0.0114	1.202	1.458	1

Figure 67.1 Selection of exchange rates for major currencies (March 2012)

(Source: http://finance.yahoo.com/currency-converter– an excellent site for up-to-the-minute exchange rates which also plots out five year trends.)

Definition: 'Exchange rate'

The **exchange rate** is the price of a currency expressed in terms of another currency. Thus the exchange rate for the US dollar can be expressed in the amount of British pounds, EURO or Mexican pesos needed to buy one dollar. (The exchange rate can also be expressed as the price of a currency against a 'basket' of other currencies.) Note that when we refer to the 'exchange rate for the dollar' we use the price of the US dollar in terms of, say, pounds. The exchange rate for the pound would naturally be expressed as the price of pounds in another currency – here, US dollars. Thus; £1 = USD1.58 and USD1 = £0.63.

Currencies are bought and sold on a *foreign exchange market.* The demand for a currency is a function of three main variables. An easy way to link to the many possible influences on the demand for a currency is to remember that it is a *derived demand* – the demand for the Australian dollar is largely derived from the demand for Australian goods and services, tourism in Australia and investment in Australia. Three main short run determinants of exchange rates exist:

1. **The demand for other countries' goods and services:** When European tourists visit the US they will need US dollars and European citizens wanting more US

goods will be met by European importers increasing their demand for US dollars in order to buy US goods. (Note that these transactions for which EU citizens need dollars correspond to the **current account.**)

2. The demand for FDI and portfolio **investment** in another country: Firms setting up subsidiaries abroad will need foreign exchange, and citizens, firms and investment houses in the US wanting to buy shares, bills/bonds or deposit money in European banks will need EUROs.

3. **Speculative** demand: When international speculators such as fund management firms and other financial institutions believe that the EURO will appreciate, they will buy EUROs in order to make a gain in selling them when the EURO rate rises.

There are thus a number of 'players' on the foreign exchange market. *People* and *export/import firms* will buy currencies from banks and foreign exchange offices to conduct international transactions. *Foreign exchange brokers* – currency traders – will in turn be used by banks to supply needed currencies and to cash-in unneeded currencies. Finally, the *central banks* of trading countries might intervene on the currency market by drawing upon (or adding to) the foreign reserves in order to adjust the exchange rate. Keep this straight; all of the players on the market can influence the price of the currency – as long as they are powerful enough!

Floating exchange rate regime

The market for a currency is just like any other market – but far more efficient and much faster. (In fact, the international foreign exchange market could well be the most perfectly competitive market of them all: abundant information via low-cost computer links; homogenous goods; low barriers to entry; and many buyers and sellers.) People, firms and importers, etc, in different countries will have a demand for foreign goods and foreign investment. Currency traders and banks will want to 'sell' as much currency as possible as they make money on the transaction via commissions – which is a service charge.[1] The foreign exchange market is simply a mechanism to put the two parties in contact with each other. Back to supply and demand.

---

1   Note that a 'No commission!' sign on a foreign exchange office simply means that there will be a larger margin between buy and sell rates. For example, an American tourist in Rome might get €75 for $US100 – but an Italian leaving for the US would get less than $US100 in exchange for €75. The foreign exchange office lives on the difference as a profit.

A currency 'floats' amongst the other currencies on the market and the price is set in accordance with the mechanisms of supply and demand outlined earlier. In short, the demand for a currency is a derived demand. It is **fundamentally** based on the demand for goods, services and investments in other countries. There is also **speculation** in currency – which is perhaps the main **short run** determinant of exchange rates – which is based on the predictions/hopes/fears of currency speculators 'betting' on changes in exchange rates.[2] We will look more closely at these determinants further on in this section.

Appreciation and depreciation

Appreciation: When the market value of a freely floating currency increases, i.e. the market price has increased, the currency has appreciated. This can be caused by an *increase in demand and/or decrease in supply* of a currency on the foreign exchange market. For example, increased demand for Indonesian goods would lead to an increase in demand for the Indonesian rupiah (IDR). If Indonesian speculators feel that the IDR is under valued, they might use it to buy US dollars which would cause an increase in the supply of the IDR.

Depreciation: When the exchange rate (expressed in terms of another currency or a trade-weighted basket of currencies) for a freely floating currency falls, then the currency has **depreciated**. *Increased supply and/or decreased demand* will cause depreciation.

Supply and demand

The supply curve for a currency is upward sloping and the demand curve is downward sloping. Using the EURO to explain why:

- **Demand:** A lower price for the EURO – *ceteris paribus* – would enable foreign citizens to buy more EURO zone goods, and foreigners would thus demand more EUROs. A lower price of the EURO is thus correlated with an increase in the quantity demanded for EUROs.

- **Supply:** The supply curve for the EURO is upward sloping since a higher price for it means EURO zone citizens can exchange their EUROs for more foreign goods. As EURO zone citizens trade in their EUROs for other currencies, the quantity supplied of the EURO increases and we get positive correlation between the price of the EURO and quantity supplied of the EURO.

I will limit the discussion to two currencies, the EURO and the US dollar, in order to make things a bit simpler and to impress upon you that since the price of the EURO is in terms of the US dollar –and vice versa – then the activities in one market will affect the other market. We thus assume that there are only *two markets* – the USD market and the EURO market.

---

2   Notice that very small changes in exchange rate margins can lead to *astounding* profits if large enough numbers are used. For example, if a speculator feels that the EURO will fall (depreciate) by 0.1% against the Swiss Franc (SFR) by tomorrow, he/she might borrow €50 million overnight and immediately buy Swiss Francs at an exchange rate of **€1 = SFR 1.202** (SFR1 = €0,832). The speculator now has SFR 60,1 million. When/ if the EURO does depreciate by 0.1% the next day, the new exchange rate will be €1 = 1.202 x 0.999, which is 1.20079 – or **SFR1 = 0,83279**. The speculator then cashes in his/her SFR60.1 million and gets €50,050,800 back. Disregarding the interest the speculator will pay on the EURO loan, an overnight profit of €50,800 is a tidy sum for five minutes work on the phone and computer terminal.

I) EURO

P of € in USD

1.35

1.31

S_0

D_1

D_0

Q_0 Q_1 Q€ (millions/day)

Initial equilibrium:
€1 = 1.31 USD
...and thus...
1USD = €0.76

New equilibrium:
€1 = 1.35 USD
...and thus...
1USD = €0.74

II) USD

P USD in €

0.76

0.74

S_0

S_1

D

Q_0 Q_1 Q$ (millions/day)

Figure 67.2 The exchange rates for the EURO and the US dollar (USD)

Diagram I in Figure 67.2 above shows how an initial equilibrium in the exchange rate for the EURO is set at USD1.31. This gives an exchange rate for the US dollar of €0.76, shown in diagram II. Say that there is increased American demand for European goods or an increase in American tourists to Europe:

- EURO: the demand for EUROs will increase from D_0 to D_1 in diagram I, and the **exchange rate for EUROs appreciates** from USD1.31 to USD1.35.
- USD: As more US dollars would flow into foreign exchange offices to buy the EURO, the supply of US dollars would increase. This is shown in diagram II, where the supply of US dollars increases from S_0 to S_1, and the **exchange rate for the US dollar depreciates** from €0.76 to €0.74. The new exchange rate for the

USD is naturally the inverse of the price of the EURO (1 / 1.35).

Please note that I use two diagrams at this stage simply to illustrate the 'price of one in terms of the other', and the shifts in demand and supply for a currency in *two markets* – the EURO market and the US dollar market. In the remainder of this chapter, I will illustrate using one currency at a time.

OUTSIDE THE BOX

Trade Weighted Exchange Rate

The exchange rate method thus far applied is in fact a bilateral exchange rate, i.e. the rate between two countries only. This neglects the total effect – the effective exchange rate – when taking into account other countries with which trade takes place. Therefore currency values are often expressed in an index, where the relative importance of trade done with various countries is taken into consideration. By putting the exchange rate in terms of a 'basket' of currencies, one can see the average change in the price of the currency. The currencies are assigned weights depending on how much trade is done with that country.

Example: Assume that Britain trades entirely with three currency areas; Euroland, the US and Japan. A trade weighted exchange rate is composed by weighting the British pound in terms of the amount of trade done with the three currencies, and indexing them over a period of time:

Currency	Trade weight	Price of £		Percentage changes in exchange rate	
		2003	2004	un-weighted	weighted
€	0.6	1.4	1.45	3.50%	2.10%
USD	0.3	1.62	1.85	14.20%	4.30%
Japan	0.1	190	200	+5.2%	0.50%
				Average: +7.6%	Trade weighted: +6.9%
					TWI 2003 = 100 TWI 2004 = 106.9

TRADE

A simple average shows that the exchange rate for the pound has increased by 7.6%, but by using weights which show the **trade patterns of Britain** the pound has increased by 6.9%. This trade weighted index (TWI above) is a far more meaningful value of the pound's rate of exchange than a simple average of the change in different currencies.

POP QUIZ 67.1

Introduction to Exchange Rates

1. The exchange rate for the US dollar is 7.8 Swedish Crowns (SEK). What is the exchange rate for the SEK?

2. What will happen to the exchange rate for the Japanese Yen if more tourists visit Japan?

3. What will be the effect on the Japanese exchange rate if Japanese imports rise?

4. Farawayistan places a large quantity of embassy personnel in Nearistan. These personnel buy lots of goods and services from Nearistani firms. Explain what will happen on the foreign exchange markets in terms of Faraway dollars and Neari pounds.

Determinants of a floating exchange rate

The exchange rate will be affected by a good many factors, and this section will deal with the main determinants of the exchange rate. Note that two important assumptions must be made:

1. The currency in question (the US dollar) is not only floating but also freely floating – there is no intervention by the US central bank;

2. Each of the determinants below is operating under the *ceteris paribus* conditions – a change in interest rates is not accompanied by a change in another variable. Note that the US dollar in Figure 67.3 is priced using a *trade weighted index* (TWI – see the previous *Outside the box*), i.e. an average price of the US dollar in terms of the main American trade partners' currencies. The

following seven examples all refer to Figure 67.3 and you must forgive me for using some economic shorthand throughout.

US dollar; change in demand
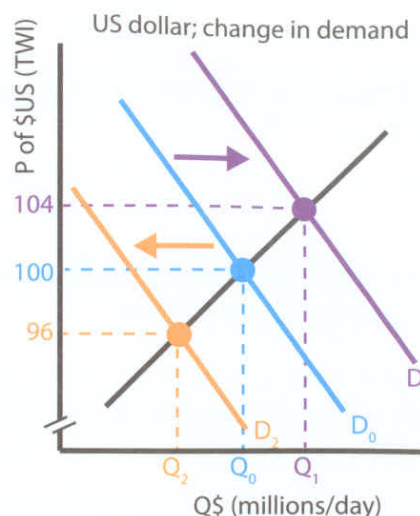
Increase in demand for the US dollar:
- $\Delta\uparrow$D for US exports of goods/services
- $\Delta\downarrow$ in US inflation rate (*ceteris paribus*)
- $\Delta\uparrow$D for foreign investment in US
- $\Delta\uparrow$Y for trade partners
- $\Delta\uparrow$ in US interest rates
- speculative buying of US dollar

Decrease in demand for the US dollar:
- $\Delta\downarrow$D for US exports of goods/services
- $\Delta\downarrow$D for foreign investment in US
- $\Delta\downarrow$Y for trade partners
- $\Delta\downarrow$ in US interest rates
- $\Delta\uparrow$ in US inflation rate (*ceteris paribus*)

US dollar; change in supply
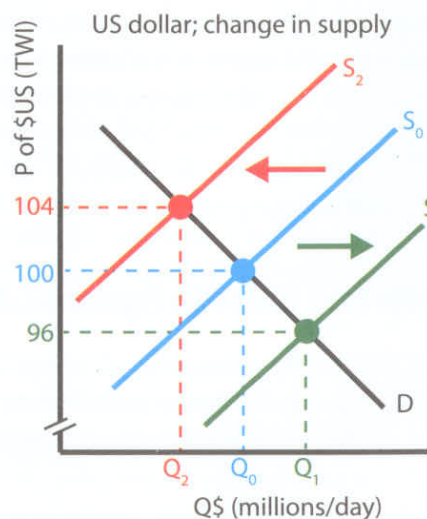
Increase in supply of the US dollar:
- $\Delta\uparrow$D in US for imports of goods/services
- $\Delta\downarrow$ in foreign inflation rate (*ceteris paribus*)
- $\Delta\uparrow$D for US foreign investment abroad
- $\Delta\uparrow$Y in US (domestic growth)
- $\Delta\uparrow$ in foreign interest rates
- speculative selling of US dollar

TRADE

475

Decrease in supply of the US dollar
- Δ↓ US imports of goods/services
- Δ↓ in US foreign investment abroad
- Δ↓ Y in US (lower US growth rate)
- Δ↓ in foreign interest rates
- Δ↑ in foreign inflation rate (*ceteris paribus*)
- speculative holding of US dollar

Figure 67.3 Determinants of the exchange rates for the US dollar (USD)

Trade flows – demand for goods and services (refer to Figure 67.4)

When American exports increase, there will be an increased demand for the dollar as importers in other countries will need more dollars to buy the American goods. The same goes for services and tourism; more banking services and tourism demanded by trade partners mean greater demand for the US dollar.

- **Increased exports** will increase the demand for the dollar, shown by the shift in the demand curve from D_0 to D_1 and subsequently an appreciation of the dollar from 100 to 104.

  o $\Delta D\uparrow$ for US exports → $\Delta\uparrow$ demand for USD → *appreciation of the USD*

- An **increase in imports** and/or and increase in American tourism abroad means that more dollars will be traded to purchase imports and buy tourist services. This **increases the supply of dollars** on the market – S_0 to S_1 – and the US dollar depreciates from 100 to 96.

  o $\Delta\uparrow$ US imports → $\Delta\uparrow$ supply of USD → *depreciation of the USD*

Relative inflation (refer to figure 67.3)

Please keep in mind our assumption of *ceteris paribus*, as it means that a change in the macro environment in the US means a *relative* change. If inflation during a time period is 5% in the US and 3% in trade partners' economies, the US relative rate of inflation has increased. The term *relative* here means that we look at the change in US goods in relation to foreign goods. If the price of US shoes rises 10% but the price of foreign shoes rises 12% then the *relative* price of US shoes has in fact fallen.

- **Increase in relative inflation:** When the rate of inflation in the US increases *relative* to that of trading partners,

US goods and services become comparatively more expensive and **US exports will fall** as will demand for the US dollar. There will also be an **increase in imports** to the US (as foreign prices are relatively lower) which means that more dollars will be traded for other currencies and the supply of the US dollar will increase – which also helps to lower the exchange rate for the dollar.

  o $\Delta\uparrow$ relative inflation in US → $\Delta\downarrow$ demand for US exports & $\Delta\uparrow$ US demand for imports → $\Delta\downarrow$ demand for US dollar & $\Delta\uparrow$ supply of US dollar → *USD depreciates*

- **Decrease in relative inflation:** In the same way, a lower relative rate of inflation in the US would therefore mean an *increase in demand for US goods* and dollars. There would also be a *decrease in US imports* and decreased supply of dollars

  o $\Delta\downarrow$ relative inflation in US → $\Delta\uparrow$ demand for US exports & $\Delta\downarrow$ US demand for imports → $\Delta\uparrow$ demand for US dollar & $\Delta\downarrow$ supply of US dollar → *USD appreciates*.

FDI and portfolio investment (refer to Figure 67.3)

One of the main reasons for the phenomenal appreciation of the US dollar during the 1990s was that it was so attractive for foreign firms to invest in the US. Strong growth, high profits and a dynamic and innovative environment attracted billions of investment Yen/EUROs/British pounds. In order to build a plant in the US or buy American shares, one must first buy US dollars; demand for dollars increases and the dollar will appreciate.

- Posit then that one or more of the following happens: profits in US companies rise; US corporate taxes are lowered; US productivity increases; or legislation is limiting foreign ownership in the US is scrapped. Any of these will affect FDI and portfolio investment from abroad positively as investment conditions and returns on investment have improved. As foreigners' demand for setting up firms and buying American shares goes up, the demand for the US dollar will increase. Hence, increased foreign investment demand in the US will cause the dollar to appreciate.

○ Δ↑inflows of investment to US Δ↑demand for USD → *US dollar appreciates*

- Conversely, when American firms set up factories abroad and US investors increase their holdings of European/Japanese bills, bonds and shares, Americans will have to buy other currencies in order to make these exchanges. The supply of the US dollar will increase, which in turn will lower the exchange rate of the dollar in countries receiving FDI/portfolio inflows.

○ Δ↑US investment abroad → Δ↑supply of USD → *US dollar depreciates*

Domestic growth (refer to Figure 67.3)

Change in national income: Interconnected and integrated countries will naturally influence each other. As has been pointed out, when incomes rise there is an increase in spending in households. Some of this spending will of course be on imports. Thus, an increase in national income in the US will increase imports and increase the supply of the US dollar.

- Δ↑Y in the US → Δ↑ imports in the US → Δ↑S of USD → *US dollar depreciates*

Growth in trade partners' economies (refer to Figure 67.3)

An increase in national income in Mexico, one of America's main trading partners, will naturally mean increased exports for the US and an increase in demand for US dollars.

- Δ↑Y in Mexico → Δ↑ demand for US exports → Δ↑D for USD → *US dollar appreciates*

Speculation (refer to Figure 67.3)

Financial investors and speculators around the globe keep trillions of Yen, US dollars, EUROs, Swiss Francs…etc as part of their portfolios. They often 'live on the edge' by buying and selling large amounts of currencies for very short periods of time – often hours. The actions of speculators are based on **expectations** of future (as in hours or days away) currency movements. Speculation is the central determinant in exchange rate fluctuations, as the 'herd behaviour' of speculators often leads to a self-fulfilling prophecy; if one ~~large wildebeest~~ international currency trader believes that the US dollar will depreciate and starts selling dollars, others might follow. What

happens then is that large scale selling of the dollar creates the very situation foreseen; a fall in the dollar due to an increase in the supply of dollars and the dollar depreciates.[3]

- Speculative belief that the US dollar will depreciate → US dollar is sold by speculators → Δ↑ supply of the US dollar → *the USD depreciates*

- Speculative belief that the US dollar will appreciate → US dollar is bought by speculators → Δ↑ demand for the US dollar → *the USD appreciates*

Case Study on Speculation:

How George Soros Cracked the British Pound in 1992[1]

1. The English pound was linked to the precursor to the EURO, via the European Exchange Rate Mechanism, ERM (a fixed exchange rate system between several EU countries), in 1992.

2. As the pound was fixed, monetary policy was limited in domestic scope since the Bank of England used the interest rate to adjust the exchange rate.

3. The bench-mark currency that all ERM countries followed was the German Deutschmark, DM. This is where asymmetry comes in – Germany was in a different position entirely from that of the rest of Europe. Europe was wallowing in low growth, high unemployment and recession (and thus desperately needed to lower interest rates). Germany had just re-unified and was spending trillions on the restructuring of Eastern Germany, which had forced up interest rates in Germany.

4. Germany feared inflation (remember the 1923 hyperinflation!) and thus raised interest rates to

---

3   This is known as a 'self-fulfilling prophecy'. When one anticipates a movement in the exchange rate and acts upon it one might trigger the same action in others. This action – here, selling the USD – then creates exactly the situation one expected. In short; when currency speculators act upon expectations they might well fuel the very forces in the direction of expectations; a self-fulfilling feedback loop is created.

TRADE

5. So, the pound was weakening vis-a-vis the DM and the British government was using currency reserves to defend the pound - e.g. using the foreign reserves to buy pounds. The currency speculators on the foreign exchange market considered the pound overvalued and there were whispers of England leaving the ERM – which of course the government vehemently denied.

The scene was thus set for a currency crisis!

6. The world's most famous currency speculator, George Soros decided not only to use the impending crisis but to create it!

7. Soros's Quantum Fund quietly established large lines of credit and borrowed $US15 billion USD worth of English pounds which he could convert into US dollars at will.

8. In essence, Soros's fund started quietly exchanging the borrowed pounds for US dollars (or other currencies). After a while this exchanging had to get NOISY!

9. Soros therefore started a high-profile media attack on the pound; he gave interviews and made other very ostentatious statements as to how the pound would soon have to be devalued – or leave the fixed exchange rate system.

10. It worked! Within weeks, the Bank of England had spent some 50 billion US dollars defending the pound but there was still massive downward pressure on the pound. By the beginning of September the Bank of England had raised interest rates drastically, but this was politically impossible to sustain in a recessionary economy. By mid-September, the pound left the ERM and was allowed to float, which it still does.

11. Soros sold back his dollars to the BoE at a lower exchange rate, paid off all his loans denominated in pounds, and made about 1 billion US dollars.

Did Soros create the devaluation?! Well, no, probably not. There was too much pressure on the pound and too many actors that considered the pound to be weak. He did, however, accelerate the issue by his actions.

Did Soros screw the Brits? Well, he did cost the British taxpayer about $US50 billion for the defence of the pound, and also humiliated the government of John Major (remember him?). However, the devaluation did not cause an economic crisis – quite the opposite, in fact. The devaluation freed the Bank of England of its shackles to the ERM and allowed interest rates to be cut to stimulate the economy. The lower exchange rate also boosted British exports. Within three years, the British economy was booming way beyond the rest of the EU zone and had amongst the lowest unemployment figures in the OECD. A happy ending, thus.

> 1 See *The return of depression economics*, Paul Krugman, pages 121 ff)

Other determinants (refer to Figure 67.3)

As mentioned in Chapter 57, one main determinant of a floating exchange rate is the level of **government deficits** and **debt** an economy has. These have a rather strong expectations-based influence on the *long run trend* of exchange rates. Large budget deficits and increasing debt in the US have made foreign investors wary, since there is a risk that the government might be forced to raise taxes and cut government spending – both of which would decrease aggregate demand and lower profits for firms, reducing inward FDI and portfolio investment and in turn lowering demand for the currency. It also created an incentive for US investors to place more of their holdings abroad, which increases the supply of the US dollar. In summa, the US dollar depreciates.

Another theory dealing with the long run determinants of exchange rates is the theory of **purchasing power parity (PPP)**. PPP exists when an identical basket of goods costs the same in two countries at a given exchange rate; at an exchange rate of USD1 = €0.75 a bundle of goods costing USD100 in the US would cost €75 in the EU since there would be perfect parity between the purchasing power of a US dollar and that of a EURO. Assuming perfect flows of goods, money and

information between trading nations, domestic consumers will detect *price differentials* and act upon them.

Say the bundle costs €70 in EMU countries at an initial exchange rate of USD1 = $0.75. Americans would get €75 and buy more goods from 'Euroland' which would drive up demand for the EURO whilst simultaneously increasing the supply of the USD. Ultimately the EURO would appreciate (and the USD depreciate) and the price differentials would be eradicated, resulting in an exchange rate where the bundle of goods is the same price in the US as it is in 'Euroland'[4]. The **theory of purchasing power** parity thus predicts that exchange rates will ultimately even-out to *reflect changes in inflation* in trading countries. In other words, in the *long run* PPP will ultimately be restored via the exchange rate.

Definition: 'Purchasing power parity (PPP) theory'

The theory of **purchasing power parity** states that exchange rates will in the long run adjust to different inflation rates in countries, leading to equilibrium where a given amount of home currency traded for a foreign currency will buy an equally-sized bundle of goods. The theory is used as a possible explanation of long run influences on the exchange rate.

However....

While there is some evidence that exchange rates in the long run show some adjustment to purchasing power, there are far too many exceptions abounding to be able to draw general conclusions. The limitations set by the assumptions of the theory would appear to be quite strong.

- Many goods are like ill-mannered infants and saké (Japanese rice wine); they simply don't travel well – e.g. services such as restaurant meals and cinema tickets are **not exportable** and many goods will have high transport costs and also meet high trade barriers.

- Goods are **not homogenous** and thus price differences will exist which show consumer preferences.

- In addition, **expenditure taxes** can vary considerably between countries and international firms will indulge in **price discrimination** whenever possible.

Point in fact, a survey of the most integrated single market in the world, the European Union, showed that prices of cinema tickets varied by 170% between lowest and highest prices in the EU.[5] Another example is Switzerland, which is surrounded by 5 countries (including Liechtenstein) and has relatively open borders towards its EU neighbours – yet it is one of the notoriously most expensive countries in the world, where a basket of goods costs an average 40%, than the OECD average. These are but a few examples of how purchasing power parity theory comes up a bit short in predicting that exchange rates will adjust and eventually even-out prices.

IN THE NEWS....

Purchasing power loonies in Canada

No, don't get upset and think I am having a go at the Canadians – I am referring to the Canadian dollar which is affectionately known as the 'loonie' due to the aquatic bird, the loon, which adorns the Canadian one-dollar coin. During 2007 the Canadian dollar appreciated by some 19% and by October it was at parity with the US dollar, e.g. CAD1 = USD1, for the first time in 31 years. This created quite a stir amongst Canadian shoppers and the concept of PPP quickly made its way into the daily vocabulary.

Canadians quickly noticed that a dollar in Canada would buy far less than a dollar in the US – for example, 'The Age of Turbulence' by former chairman of the Fed Alan Greenspan had a price of CAD26.45 in Canada but only USD20.99 in the US. A Toyota Avalon was priced at CAD41,000 in Ottawa but USD31,000 just an hour's drive south of the US border. The Bank of Montreal calculated that the price of a random basket of goods was an average of 24% higher in Canada than in the US.

---

4   In this case the €70 bundle must equal the USD100 bundle in price at the new exchange rate. This implies a PPP exchange rate of USD1 = €0.70 which consequently means that the *EURO has appreciated* from 1.33 to 1.43 dollars per EURO.

5   *The flaw of one price*, The Economist, Oct 16th 2003

The strong Canadian dollar had immediate effects on shopping trips, where Canadians would head south and stock up on goods in the US. Naturally the increased imports hit domestic Canadian merchants hard and there were even attempts by large automobile manufacturers to bar the selling of cars in the US to Canadian citizens. Canadian exporters also felt the results of the higher exchange rate as foreign sales fell and profits repatriated from subsidiaries abroad meant less when converted to Canadian dollars.

Was the rise in the Canadian dollar justified or 'too high'? Good question. According to the OECD, the PPP value of the Canadian dollar should have been 81 US cents. Thus, the loonie was overvalued by some 23%. Was the outcome in line with PPP theory? Well, yes; after hovering at around CAD1 to USD1 between November 2007 and August 2008, the loonie started to fall…and then fell like a paralysed, well, loon. By November of 2008, the Canadian dollar was at the estimated PPP value of 80 US cents to the Canadian dollar and has fluctuated around this level up to the time of writing.

However, there are a goodly number of other variables which have affected the exchange rate – primarily the fact that the US dollar has noticeably strengthened during the global financial/ credit crisis of 2008. Two reasons stand out: 1) As the US economy reeled in recession with rising unemployment and falling GDP, imports decreased drastically. This decreased the supply of US dollars on the foreign exchange market. 2) International capital is 'cowardly' and in time of crisis it will do like a loon in a storm out on the lake; seek shelter on the biggest rock possible. The US is a big rock. When investors/ speculators seek shelter they buy US government securities and this increases the demand for the US dollar.

(Sources; "*Canadian shoppers up in arms as "loonie" prices fail to reach parity*"; Financial Times Oct 2 2007. OECD estimates of PPP for Canada, see http://www.td.com/economics/special/ cb0907_parity.jsp)

POP QUIZ 67.2

Exchange Rate Determinants

1. European tastes switch in favour of East Asian culture and goods. What would happen on the foreign exchange market for the EURO?

2. Say that Britain grants 'tax holidays' for foreign firms. How would this affect the exchange rate for the British pound?

3. Korea manages to keep its inflation rate lower than that of its trading partners. How might this affect the Korean Won?

4. The European Central Bank (ECB) considers the exchange rate for the EURO to be too low. What actions can be taken?

5. Here's a tricky one: Using a S/D diagram for the Swiss Franc, show how the repatriation of profits to Switzerland by Swiss firms in Germany would affect the value of the Swiss Franc.

Effects of the exchange rate on macro goals

This is economics so the odds are fair that any two variables will show causal flows in both directions. This is indeed the case where currency depreciation/appreciation will affect domestic growth, inflation, unemployment and the balance of payments.

Domestic AD and GDP

For any economy with a degree of openness to trade, the exchange rate will have an effect on GDP via exports and imports. A currency depreciation will mean that exports are (relatively) cheaper and imports (relatively) more expensive – a resulting increase in export revenue and decrease in import expenditure would stimulate AD and increase national income.

- Depreciation of the USD →

  o $\Delta\downarrow$ *relative* P of exports in the US $\rightarrow \Delta\uparrow$ US export revenue $\rightarrow \Delta\uparrow$ AD

  o $\Delta\uparrow$ *relative* P of imports in the US $\rightarrow \Delta\downarrow$ US import expenditure $\rightarrow \Delta\uparrow$ AD

Inflation

Consider what might happen in the US if the exchange rate for the US dollar were to appreciate. This would cause households and firms to switch some of their expenditure to imports and this have a dampening effect on AD in the US.

TRADE

- Appreciation of the USD →

  - → Δ↓ *relative* P imports in the US → Δ↑ US imports → Δ↓AD → Δ↓ inflation rate

  - US firms import a major portion of factors → Δ↓P imported factors → Δ↑AS → Δ↓ inflation rate

  - It is also likely that the increased competition with foreign suppliers will force US firms to become more price competitive and this too will have a dampening effect on US inflation.

Unemployment

As a continuation of the previous examples, a weaker or stronger US dollar which has a stimulatory/dampening effect on AD might have repercussions on unemployment.

- Depreciation of the USD → Δ↑ exports in the US → Δ↑AD → Δ↓ unemployment

- Appreciation of the USD and thus Δ↓ exports → Δ↓AD in the US → Δ↑ unemployment

Balance of payments

We will deal with the balance of payments in some depth in Chapters 70 to 73 so for the time being we shall limit the iteration to two brief examples of how a change in the exchange rate can have an impact on net exports in current account.

Current account

Recall that current account is basically an account of all in- and out-flows of money to/from an economy arising from trade in goods and services.

- *Depreciation* of the USD →

  - → Δ↓ (relative) P of US exports → Δ↑ US export revenue → *improvement* of current account balance

  - → Δ↑ (relative) P of imports in the US → Δ↓ US import expenditure → *improvement* of current account balance

- *Appreciation* of the USD →

  - → Δ↑ (relative) P of US exports → Δ↓ US export revenue → *worsening* of current account balance

  - → Δ↓ (relative) P of US imports → Δ↑ US import expenditure → *worsening* of current account

Government debt

While not directly one of the main macro objectives, it is clearly desirable for countries to ultimately balance the budget. A government which has taken out foreign loans will have to service this debt (i.e. pay interest and repay the principle loan) in the foreign currency of the creditor nation. A depreciation of the Home currency means an increased debt burden in terms of the foreign debt element of overall government debt.

TRADE

Summary & revision

1. The **exchange rate** is defined as the *price of a currency in terms of another* – or the price of a currency in terms of a (trade weighted) basket of other currencies.

2. **Demand** for a currency is largely a form of *derived* demand. The demand for the € is determined by the underlying demand for goods/services from the EURO zone, investment in the EURO zone and speculation in the EURO.

3. A **floating exchange rate** regime means that the exchange rate is determined by the market forces of supply and demand.

4. In a floating exchange rate regime, a fall in the price of a currency is called *depreciation* while an increase in the price of currency is called *appreciation*.

5. The **demand for a currency** shifts due to

   a. A change in *export demand*

   b. A change in *relative inflation*

   c. A change in demand for foreign *inward investment*

   d. A change in foreign *trade partners' income*

   e. A change in domestic *interest rates*

   f. *Speculative buying* of the currency

6. The **supply of a currency** shifts due to

   a. A change in *demand for imports*

   b. A change in *relative inflation*

   c. A change in *domestic income*

   d. A change in *foreign investment abroad*

   e. A change in *foreign interest rates*

   f. *Speculative selling* of the currency

7. **Other determinants** of the exchange rate include:

   a. *PPP theory* – e.g. that exchange rates in the long run will run inversely to inflation rates since higher domestic inflation will lead to increased imports and hence an increase in supply of the domestic currency.

   b. Government *deficits and debt*; there will be strong expectations-based speculation that deficits and debt ultimately weaken an economy. This decreases the demand for the currency of the debtor nation.

8. A **depreciation/appreciation** can in turn affect:

   a. *Domestic GDP* (a depreciation can result in increased export revenue)

   b. *Inflation* (increased exports stimulate AD which can result in inflation)

   c. The *current account* (increased export revenue is accounted for as an inflow in current account)

   d. *Unemployment* (increased exports which stimulate AD can result in lower unemployment)

   e. Government debt servicing (a depreciation of the home currency has the effect of increasing the cost of foreign debt servicing)

68. Exchange Rate Calculations

Key concepts: HL extensions

- Exchange rate calculation
- Supply and demand – linear functions
- Price of goods in another currency
- Using currency data

HL students will be required to do some simple computations on exchange rates. A few minutes revision of the HL sections in Chapters 4 to 6 is time well spent. Calculations and answers are at the end of the chapter summary.

Exchange rate calculation

The matrix below gives three different currencies and the US dollar's value against them.

Currency April 2012	Japanese Yen (JPY) to the dollar	Thai Baht (THB) to the dollar	Norwegian Krone (NOK) to the dollar
U.S. Dollar (USD)	80	30	6

1. What is the exchange rate for the NOK in terms of the THB?

2. What is the exchange rate for the THB in terms of the JPY?

Supply and demand – linear functions

Assuming a freely floating exchange rate regime and 100% convertibility between the USD and the Japanese Yen, the demand and supply functions for the USD are given as:

$QD_{(USD)} = 2,400 - 15P_{(ex)}$

$Qs_{(USD)} = -800 + 25P_{(ex)}$

3. Calculate equilibrium exchange rate.

4. Draw a supply and demand diagram illustrating market equilibrium for the USD.

5. American speculators feel that the JPY is undervalued and there is speculative buying of the JPY via American dollars. The relevant curve shifts 20% at all price levels. Calculate the new equilibrium exchange rate for the USD (expressed in JPY).

6. After a few weeks at the new equilibrium, the US central bank (Fed) raises interest rates. This causes a shift in the relevant curve by 10% at all price levels. Calculate the new equilibrium exchange rate for the USD.

Price of goods in another currency

My lady Bell finally gets hold of the Seiko Orange Monster automatic divers watch for me! It costs JPY18,500 in Kyoto, NOK2,400 in Oslo and THB7,200 in Bangkok.

7. Where should she buy my new watch? Oh, yes, she gets paid in USD.

8. Should she order the watch and have it sent to our home address in Indonesia?

Using currency data

Figure 68.1 shows how the Swiss Franc (CHF) compared to the EURO (€) has developed over a five year period. Figure 68.2 shows the Australian dollar (AUD) vs the EURO during the same period.

Summary & revision
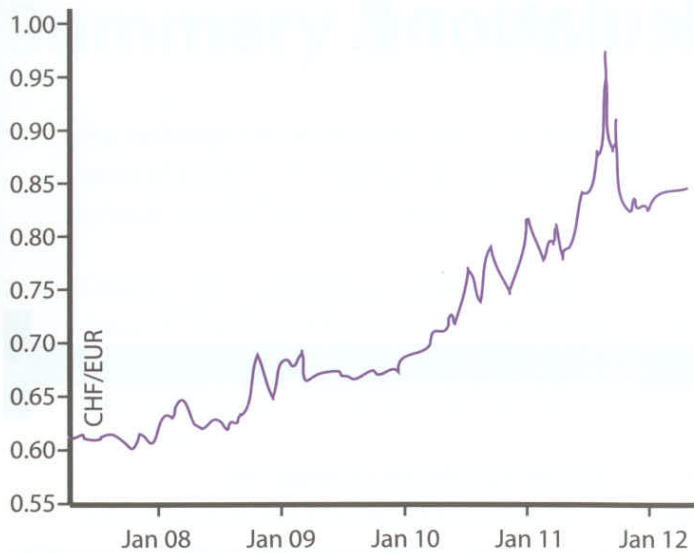
Figure 68.1 Price of the CHF in €
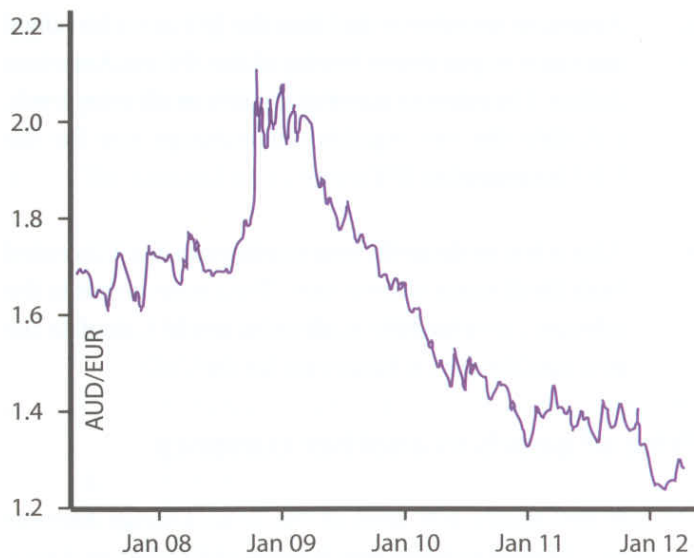
Figure 68.2 Price of the € in AUD

9. How has the EURO done during this period of time – has it become stronger or weaker?

10. How has the exchange rate for the AUD evolved in terms of the EURO?

11. What say you about the exchange rate for the CHF in terms of the AUD?

12. If you had a time machine and could go back to January 2008, would you have put your savings into Francs or Australian dollars?

1. The NOK is trading at THB5 to the NOK. $\left(\dfrac{THB30}{NOK6}\right)$

2. The THB is trading at JPY2.667 to the NOK. $\left(\dfrac{JPY80}{THB30}\right)$

3. Equilibrium (solving for price); $2,400 - 15P_{(ex)} = -800 + 25P_{(ex)}$

$3,200 - 15P_{(ex)} = 25P_{(ex)}$

$3,200 = 40P_{(ex)}$

$P_{USD} = JPY80$

Solving for quantity; $2,400 - 15 \times 80$ or $-800 + 25 \times 80$

$2,400 - 15 \times 80 = 1,200$

4.
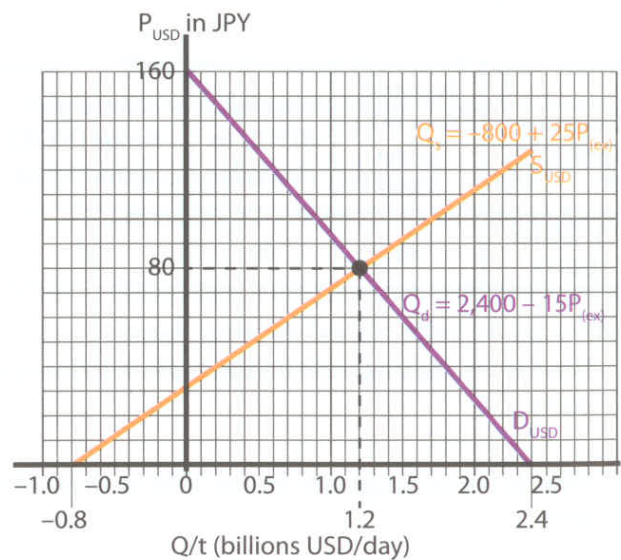
5. Speculative buying by US speculators means an *increase in the supply of the USD*. The new supply function is $-640 + 25P_{(ex)}$ (i.e. the Q-axis intercept for the S-curve has increased by 20%). New equilibrium means solving $P_{(ex)}$ in $2,400 - 15P_{(ex)} = -640 + 25P_{(ex)}$

$3,040 - 15P_{(ex)} = 25P_{(ex)}$

$3,040 = 40P_{(ex)}$

$P_{USD} = JPY76$

6. When the Fed raises the rate of interest there will be speculative inflows to the US. This increases the demand for the USD. The new demand function is $Q_{D(USD)} = 2,640 - 15P_{(ex)}$ (an increase in the Q-axis intercept by 10%). New equilibrium price is $2,640 - 15P_{(ex)} = -640 + 25P_{(ex)}$.

$3,280 - 15P_{(ex)} = 25P_{(ex)}$

$3,280 = 40P_{(ex)}$

$P_{USD} = JPY82$

7. In the US quite frankly since it is a most competitive market and far cheaper than the three options given. However, the USD price is USD231.25 in Japan; USD400 in Norway; and USD240 in Bangkok.

8. You kidding?! Revise Chapter 65.

9. The EURO has depreciated against both the AUD and the CHF.

10. The AUD has of course appreciated against the EURO.

11. The exchange rate for the CHF in EUROS went from CHF1 = 0.6 € in January 2008 and CHF1 = 0.85€ for the SFR in Jan 12. That means that the CHF has appreciated against the EURO by 14%.

12. In January 2008 the values were: €0.6 = CHF1…so €1 = CHF1.667. The Australian dollar was at AUD1.6 = €1. Thus AUD1.6 = CHF1.667. The exchange rate for the Swiss Franc was:

$$CHF1 = AUD0.96 \left(\frac{AUD1.6}{CHF1.667} \right)$$

In January 2012 the situation was: €0.85 = CHF1…so €1 = CHF1.17. AUD1.25 = €1. Thus AUD1.25 = CHF1.17. The exchange rate for the Swiss Franc was CHF1 = AUD1.068 (AUD1.25/CHF1.17).

In summa; one should have put one's savings in the Swiss Franc. Oh yes, I just remembered; I did! Nothing like a global economic crises to raise the value of 'safe-haven' currencies. Let me stick my neck out a bit further while I'm at it. If you are reading this during 2012 I strongly recommend you buy the Chinese Yuan.

TRADE

69. Government Intervention – Fixed and Managed Exchange Rates

Key concepts:

- Fixed or pegged exchange rate
- Devaluation and revaluation
- Managed exchange rate
- Over- and under-valued currencies (advantages/disadvantages)
- Advantages and disadvantages of fixed and floating exchange rate regimes

Exchange rates are seldom as straightforward as the price of a currency being set purely by market supply and demand. For various historical, political and economic reasons, governments have at times linked currencies in a fixed exchange rate system.

Fixed/pegged exchange rate

In the foreign exchange market examples used thus far, we have assumed that the exchange rate is established by market forces of supply and demand. When a group of two or more countries instead intentionally keep their exchange rate constant, a **fixed exchange rate** has been established. Typically today, countries wishing to establish a fixed exchange rates regime do so by pegging their own currency to another or to a basket of currencies. The central bank then intervenes on the foreign exchange market in order to keep the exchange rate within a narrow band.

The central bank can affect the exchange rate in the **short run** by *buying or selling its own currency* on the foreign exchange market (foreign exchange intervention – Forex intervention), and by adjusting the *interest rate* to influence investors'/speculators' demand for the currency.

In the **long run**, governments might intervene using *fiscal policies, supply-side measures and protectionism* to adjust national income in order to increase or decrease exports and/or citizens' propensity to import. Changes in exports would affect the demand for the home currency while a change in imports would affect the supply of the home currency. (See Chapter 68.)

> **Definition: 'Fixed/pegged exchange rate'**
>
> A **fixed exchange rate** links a currency to another (or basket of other currencies) within a narrow band, or 'corridor'. The central bank keeps the currency fixed in the short run by *buying/selling its own currency* on the foreign exchange market, and by *increasing/decreasing the interest rate*.

Devaluation and revaluation

When a currency in a fixed or pegged regime is realigned by the government to a lower exchange rate, the currency has been **devalued**. For example, when Venezuela readjusted the Venezuelan Bolivar in February 2004, amidst a degree of fiscal turmoil, the Bolivar fell from a peg of 1,598 Bolivars per US dollar to 1,918 per dollar – which means the exchange rate for the Bolivar went from USD0.00062 to USD0.00052 per Bolivar. This is a devaluation of 16.1%.

Naturally, a government which is operating a pegged exchange rate regime can re-peg its currency at a higher exchange rate, in which case the currency has been **revalued**. This was the case in aforementioned Venezuela in February 2003, when the case was exactly the opposite; the Bolivar was adjusted upwards (re-pegged) from USD0.00052 to USD0.00062 per Bolivar, a revaluation of 19%.

Definition: 'Devaluation' and 'revaluation'

When the government/central bank of a country running a fixed or pegged currency realigns the exchange rate to a lower value, the currency has been **devalued**. When a fixed or pegged exchange rate regime realigns a fixed/pegged currency at a higher rate, the currency has been **revalued**.

How a pegged exchange rate works

In the short to medium term, a pegged exchange regime is upheld by keeping the exchange rate within a narrow band by central bank intervention. Using the Danish krone (Danish crown in fact; DKK) as an example, Diagrams I and II in Figure 69.1, illustrate how Denmark's central bank, Nationalbank, pegs the DKK to the EURO at an exchange rate which is allowed to fluctuate in a 2.25 percent band around a target rate of €0.134 to the DKK – a floor of €0.1309 and a ceiling of €0.1370 to the DKK.

'It's pegged – that's what the hole is for!'

In order to maintain the fixed exchange rate, the Danish Nationalbank would use the *rate of interest* and the *Forex intervention* to keep the peg to the EURO:

- Keeping the krone down: Diagram I shows how the long run goal of an exchange rate of €0.134 to the EURO is attempted. At point A the ceiling is broken and the Nationalbank can do one or both of the following:

  - **Sell DKK (= buy EUROs):** When the Nationalbank buys EUROs (or indeed any other foreign currency) it uses the DKK. This intervention selling of the DKK increases the supply of the Danish krone on the Forex market from S_{DKK_0} to S_{DKK_1} moving the exchange rate for the DKK back down to €0.1370 at **point B**.

  - **Lower the interest rate:** If the central bank decreases the interest rate, the demand for the home currency would decrease (D_{DKK_1} to D_{DKK_2}) as foreign investors/speculators would see a lower *rate of return* in holding the currency. The exchange rate falls to €0.1370 at **point C**.

  - In extreme cases the Nationalbank could both sell the DKK and lower the rate of interest. This is illustrated by an increase in supply and decrease in demand for the Danish krone, resulting in the long run target rate of €0.1340 to the DKK.

I: Keeping the DKK down
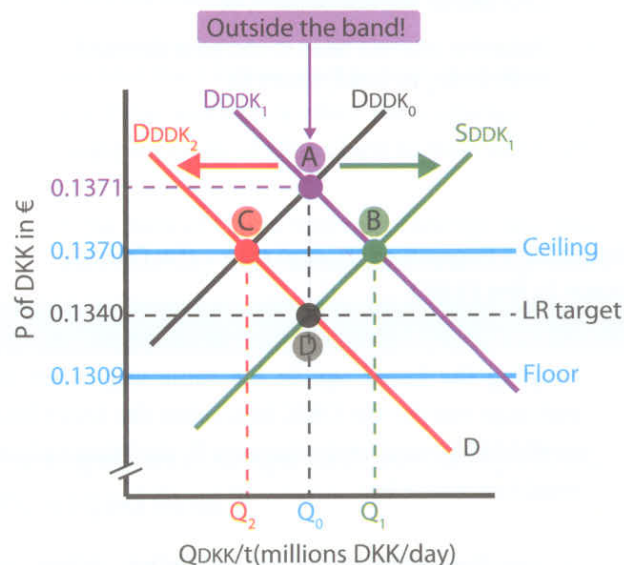
If the exchange rate for the DKK rises above €0.1371 (point A) the Nationalbank can:

- Sell $Q_0 \Leftrightarrow Q_1$ DKK (S_{DKK_0} shifts right to S_{DKK_1}) on the foreign exchange market to keep the DKK within the band (point B)

- Lower the interest rates to decrease demand from D_{DKK_1} to D_{DKK_2} (point C)

- A mixture of both which brings the exchange rate for the DKK down to the target of DKK0.1340 to the EURO at point D

TRADE

II: Keeping the DKK up
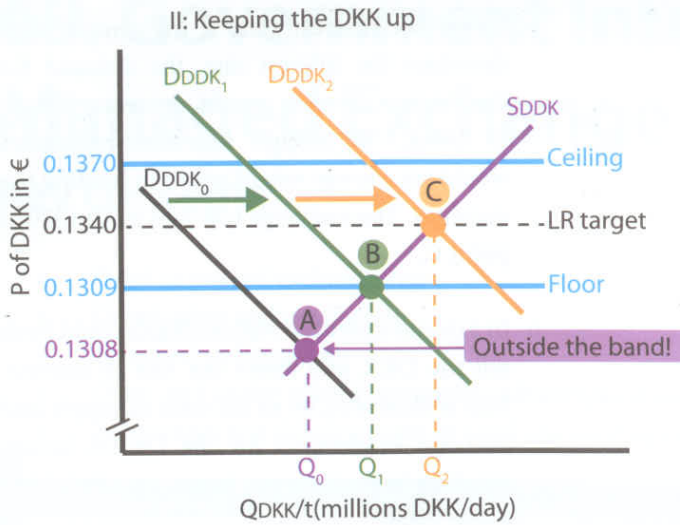
If the exchange rate for the DKK falls below €0.1309 (point A) the Nationalbank can:

- Buy $Q_0 \Leftrightarrow Q_1$ DKK (D_{DKK_0} shifts right to D_{DKK_1}) on the foreign exchange market to keep the DKK within the band (point B)

- Raise the interest rates to increase demand from D_{DKK_0} to D_{DKK_1} (point B)

- A mixture of both to bring the exchange rate for the DKK up to the target of DKK0.1340 to the EURO at point C

Figure 69.1 Pegged exchange rate – the Danish krone to the EURO

- **Keeping the krone up**: In the same way, when the exchange rate for the DKK falls below the lower limit, to €0.1308 at point A in diagram II, the Nationalbank would intervene by:

  o **Buy DKK (= sell EUROs)**: When the Nationalbank buys its own currency back it uses the EURO or any other foreign currency. This intervention buying of the DKK increases demand for the Danish krone on the Forex market from D_{DKK_0} to D_{DKK_1} causing the exchange rate for the DKK to increase to €0.1309 at **point B**.

  o **Raise the interest rate**: If the Nationalbank raises the interest rate, foreign investors/speculators would see a higher *rate of return* in holding the DKK and the demand for the Danish krone would increase (D_{DKK_0} to D_{DKK_1}, point B again).

o In extreme cases the Nationalbank could both buy the DKK and increase the rate of interest, illustrated by a 'double' increase in demand (D_{DKK_0} to D_{DKK_1} to D_{DKK_2}) for the Danish krone, whereby the DKK rises to the long run target rate of €0.1340 at **point C**.

Does it work? Have a look at Figure 69.2 and have a think. The answer would seem to be "Very well!" The DKK has stayed within a very narrow limit over five years – far narrower than the +/- 2.25% allowance set by the Danish National Bank.
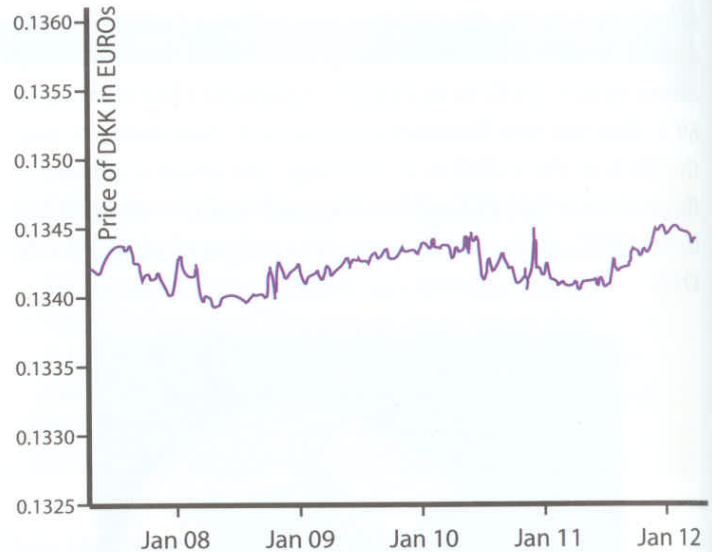
Figure 69.2 Pegged exchange rate – the Danish krone to the EURO
(Source: www.finance.yahoo.com/currencyconverter

However ...

There is a price tag attached to setting interest rates in order to keep a currency's exchange rate stable, namely the trade-off in not simultaneously being able to use monetary policy for other macro objectives. (This is covered further on.) During the onset of the financial crises in 2008, the Danish Nationalbank raised interest rates to a high of 5.5% in trying to protect the DKK from a massive sell-off (primarily to the CHF) and was forced to spend some DKK55 billion (USD10 billion). This obviously inflicts costs on aggregate demand, and unemployment – not to mention the opportunity costs of USD10 billion that could have been used otherwise. The head of the IMF in Denmark was quoted as saying "…this type of issue can be quite costly

TRADE

for the central bank without any clear benefits..." and that "... having been in the EURO would have made it easier for the Danes to deal with the crisis."[1]

A Little Depth

Other exchange rate tools available to central banks

Central banks are not limited to buying and selling their currencies in order to influence the exchange rate. There are two other alternatives in the short run:

The central bank could borrow from the **International Monetary Fund**, the IMF (see Section 5.5). The IMF was created at the Bretton Woods conference for precisely this purpose; to aid countries having difficulty in keeping a stable exchange rate. When a country's currency falls, and the central bank runs out of foreign exchange to buy up the home currency, the central bank can **borrow funds from the IMF** to get over the crisis.

In addition to the short run exchange rate policies above, a country might have to resort to more **fundamental adjustments** if the home currency showed signs of long run weakness. The currency could be re-pegged at a lower exchange rate; the government could implement **deflationary policies** to reduce import demand and thus the demand for other currencies; **protectionist measures** could also lower imports; and various **supply-side policies** could increase the long run competitiveness of the economy and thereby increase demand for the home currency. See Chapter 72; **expenditure-switching** and **expenditure-reducing** policies.

POP QUIZ 69.1

Fixed Exchange Rates

1. The Chinese government pegs the Yuan to the US dollar at a rate of 6.3 Yuan = 1 dollar: What is the exchange rate for the Yuan?

2. Assume that the Bank of China will allow a fluctuation of the Yuan of 2% up or down. Between which values will the Yuan be allowed to move?

3. The Yuan becomes stronger on the market and looks like exceeding the ceiling of the peg. What does the Bank of China do?

4. How does the Bank of China's action in question 3 above affect the balance of payments in China?

5. Say that the Chinese start to import massively from the US. How would this affect the exchange rate and how might the Bank of China have to react ultimately?

6. If the Bank of China runs out of foreign reserves, is there any other way it can influence the exchange rate?

Managed exchange rate (or 'dirty float', 'managed float')

Completely fixed/pegged exchange rates create a problem of **inflexibility**, since economies will differ over time in fundamentals such as growth rates and inflation rates. In the long run, it could be very costly for a country with a weakening currency to defend its link to other currencies, for example by running down the foreign reserves and dampening a recessionary domestic economy by raising interest.

For this reason, fixed/pegged exchange rate regimes are largely a thing of the past[2]. The most common form of exchange rate regime is a floating exchange rate system (outlined earlier) where the price of a currency is determined by the forces of

2 There are some 17 currencies currently pegged to the US dollar – for example the Jordanian dinar, Bahrain dinar, Omani rial, Quatari rial, Saudi rial, Emirati dirham and the Venezuelan Bolivar. Why do you think these currencies are pegged to the US dollar? Ask yourself what the common denominator might be between these nations.

TRADE

supply and demand. In between the two 'extremes' of a floating and pegged regime, one finds a **managed exchange rate regime**. A managed float means that the government and central bank allow the currency to float freely but will intervene at times to keep the currency within given boundaries.

> **Definition: 'Managed and pegged exchange rate system'**
>
> When a country allows a currency to float freely but intervenes in order to adjust the exchange rate, one speaks of a managed exchange rate regime. It is sometimes referred to as a 'dirty float'. An example is the Singapore dollar.

A managed exchange rate is a bit like trying to be 'a little bit pregnant', i.e. not quite floating or pegged. The advantage is that the exchange rate is *reasonably predictable* over time and this encourages firms and households to conduct foreign trade. It also allows a certain *freedom of movement* for the government in setting domestic monetary policy as the currency is allowed to fluctuate 'within reason'.

Managed exchange rate regimes; a few examples

The Singapore dollar (SGD) has been managed against an undisclosed basket of currencies since 1981. The central bank of Singapore, the Monetary Authority of Singapore (MAS), allows the SGD to float within a 'band' and allows the exchange rate to move up and down within this band. The MAS uses the same policy tools outlined earlier; buying and selling SGD and adjusting the interest rate. (Refer back to Figure 69.1– a pegged exchange rate for diagrammatic illustration of how central bank intervention affects supply and demand for a currency.)

It is worth mentioning that few countries have a *completely* floating exchange rate. There are strong political and economic reasons for a government to intervene on the foreign exchange market in order to bolster or weaken the exchange rate. For example, during October 2008 the Mexican central bank sold off some USD15 billion in order to halt a falling peso exchange rate – thousands of upper-middle class households basically fled the MXN and bought US dollars during the impending financial crisis. The same fear factor created a huge upswing in demand for traditionally-seen 'rock solid' currencies such as the Swiss franc. The caused the Swiss central bank to intervene very heavily during the spring of 2009 to keep the value of the Swiss Franc (CHF) down and thus remain competitive

internationally. The Swiss National Bank bought billions of EUROs and lowered the interest rate to 0.25% from 2.75% (a decrease of 91%) and yet the CHF still appreciated by over 15% by July of 2009.[3]

Most countries will at one point intervene in order to stabilise the exchange rate. In fact, perhaps the only major currency which has not been subjected to interventionist Forex operations by the home country central bank during the past 30 years is the US dollar – this policy of 'benign neglect' has been a cornerstone of US exchange rate policy since the early 1970s under President Nixon.

China; fish or fowl?

It becomes rather hazy to clearly define and outline the exchange rate regime currently in place in China. It has been pegged during various long intervals and then managed. Whatever the case, in both a pegged and managed exchange rate regimes there is often an incentive for the government and central bank to keep the currency undervalued. This criticism has been levelled at China for most of the first decade of the 2000s, primarily by the US which has the world's largest current account deficit. In keeping the Yuan (or Renminbi) lower than what market forces would set, China makes its export goods more competitive.
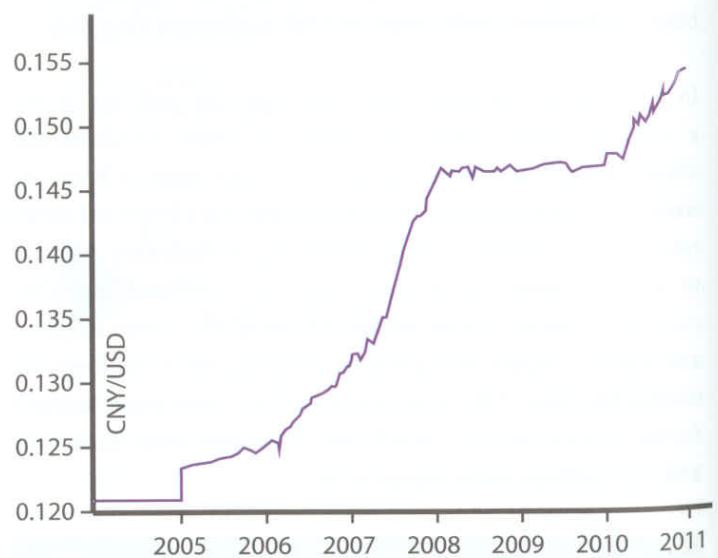

Figure 69.3 The flight of the Yuan, 2005 – 2011

(Source: http://www.forexblog.org)

---

3   Business Week, 9 – 15 Aug 2010, *How the Swiss curbed the uber-Franc.*

The Yuan was pegged at CNY8.3 to the USD (CNY0.1204 to the USD) and fluctuated very little between 1995 and 2005 – less than 0.2%. During the first years of 2000 there was increasing international pressure on China to **revalue** the Yuan in order for major trade countries (read; USA) to be able to compete with Chinese goods. In 2005 China dropped the peg to the US dollar but adopted an 'adjusting peg' towards a wider basket of currencies. By 2008, the Yuan had risen to CNY0.1460 to the US dollar, an appreciation of over 20% and was again pegged – this time at a rate of CNY0.1464 to the dollar. During 2010, China once again buckled under to very loud international pressure and since then has allowed a 'managed appreciation' (for lack of a better term) of the Yuan. Figure 69.3 illustrates this. There are a few notable long term costs associated with this foreign exchange policy:

- An artificially low currency means that the central bank must continuously intervene by way of keeping interest rates low and by selling the domestic currency (= buying another currency – in China's case the US dollar). The effect of this is to severely restrict the central bank's ability to set interest rates aimed at domestic economic goals such as taming *inflation*. This has indeed been the case in China, where average inflation between 1994 and 2010 was 4.25%.[4]

- A low exchange rate might increase export competitiveness but the *'comfort zone'* of an undervalued currency dissuades domestic innovation and productivity increases in the long run. In fact, the deputy governor of the People's Bank of China (China's central bank) said as much in late 2010; "Adopting a more flexible exchange-rate regime serves China's long-term interests as the benefits ... far exceed the cost in reorganising industries and removing outdated capacities." [5]

Story-Time!
Stalinist Exchange Rate Mechanism

A popular story amongst historians tells of how the Minister of Finance in the Soviet Union was asked by Stalin to advise him in setting an exchange rate for the rouble. The Minister had a thick file which he leafed through during his meeting as he put forward his case for an exchange rate of $US1 to 14 roubles. This apparently displeased Stalin as it made the rouble look inferior, so he leaned over and drew a line through the '1' in the figure '14'. This established the exchange rate of the rouble at $US1 to 4 roubles.

Over- and under-valued currencies

Exchange rate is 'too high'

It is clear that governments often intervene on the foreign exchange market in order to adjust the exchange rate. A question that arises is of course "Why?" One often reads about how the exchange rate is 'too high' or 'too low' and that domestic firms and/or governments feel that intervention is justified or necessary. Exchange rates clearly have numerous economic consequences and the iteration below brings up some of the more evident effects on an economy. Please note that the headings below have quotation marks around 'Too high' and 'Too low' – the reason is that there is no absolute definition of the *correct* exchange rate.

Positive effects of 'overvalued' currency:

- *Importers* will win since they will pay less for imported goods and receive their revenues in domestic currency. This could also well have beneficial effects on the domestic *inflation* rate since competitive forces will cause importers to pass on some of their cost savings to consumers.

- *Firms* which rely to a large extent on imported raw materials, components and other factors of production will see their costs go down. This can in fact make them more competitive on the international market.

- *Households* will see that imports and trips abroad will be cheaper in terms of the quantity of domestic

TRADE

---

4   www.tradingeconomics.com/china/inflation-cpi
5   Economist, August 19 2010; *Wiggle it. Just a little bit.*

currency needed to buy imports and travel abroad. Imports in fact add considerably to the standard of living.

- A *government* which has a large external (foreign) debt would see that the debt servicing (amortisation and interest payments) would be easier.

- *Export firms* in a country with a strong currency might be forced to become more *efficient* in order to be able to compete on international markets.

Negative effects of 'overvalued' currency:

- The *current account* in the balance of payments can be adversely affected if the strong currency leads to increased import expenditure and/or decreased export revenue. There has been clear correlation between a stronger dollar and the US current account deficit.

- A strongly export-orientated country, for example China, would fear the effects of a higher exchange rate on *growth* and *unemployment*. The official figures show that exports account for 37% of GDP and one can well imagine the effects on domestic growth and unemployment due to a strong currency.[6]

- Domestic firms which have large *foreign investments* will see that profits decline when they are repatriated.

Exchange rate is 'too low'

Positive effects of "undervalued" currency:

- Naturally the positive effects of an 'undervalued' currency are something of a 'mirror image' of the negative effects given above: A weaker currency lowers the price of exports and will benefit exporters and domestic industry; as the price of exports has fallen and the price of imports has risen there will be a positive effect on the balance of payments (HL: see the Marshall-Lerner and J-curve in Chapter 73); jobs will be created in export industries; countries with a high proportion of exports in relation to GDP will see growth; and international firms repatriating profits will enjoy greater returns in terms of domestic currency.

6 The official figure of '37% of GDP' should be taken with scepticism. There are some serious studies showing that the official figures are seriously inflated and the real value lies closer to 10%.

Negative effects of 'undervalued' currency:

- An 'undervalued' currency will have the following effects: Importers and firms reliant on imported factors of production will see costs rise; households will cut back on imported goods and foreign travel; and governments with a high proportion of foreign debt will see how debt servicing becomes dearer.

- A country with a 'low' or 'undervalued' exchange rate will in all likelihood experience *inflation*. There are two contributing causes:

  - A strong exporting nation will experience rising aggregate demand and *demand-pull* inflation.

  - A nation with a high volume of (HL: relatively price inelastic) imported factors of production will see increased expenditure on imported factors. This will decrease aggregate supply and thus contribute to *cost-push* inflation.

Advantages and disadvantages of fixed and floating exchange rate regimes

Like in so many other areas in economics, there are a number of trade-offs in choosing a certain exchange rate regime over another. Experience tells us that in choosing fixed over floating exchange rates, countries will both gain and lose. There is no perfect system of exchange rates.

Advantages of fixed exchange rate regimes

- **Predictability and certainty:** Fixed exchange rates make it easier for importers and exporters to calculate earnings. Costs, revenues and profit margins are clear and *predictable*. This creates an incentive for firms to invest and households to engage in entrepreneurial activity.

- **Exchange rate stability encourages trade:** When exporters and importers can be sure of tomorrow's exchange rate and future profits, it is easier to plan business. International stability in exchange rates lowers barriers to partaking in international trade and thus *increases trade*. However, there is limited merit in the argument that stabile exchange rates create trade, as the increase in global trade has in fact increased faster

since the breakdown of the Bretton Woods exchange rate system in 1971/72.

- **Fiscal/monetary discipline domestically:** This has been one of the main arguments in favour of fixed exchange rates. Recall that inflation and government deficits exert downward pressure on the exchange rate. Since the central bank will have to use limited foreign reserves to adjust the exchange rate, governments in fixed exchange rate regimes will keep inflation and deficits to a minimum in order to avoid depleting the reserves. This *limits expansionary fiscal and monetary policies.*

  o Another aspect of the discipline issue is that countries within a fixed exchange rate system will not be able to run up large current account deficits in the long term. Stimulating aggregate demand leads to increased imports and inflation – both of which will create net outflows in current account. A current account deficit means that more domestic currency is hitting the international market which will exert a downward force on the exchange rate.

- **Less risk of speculation:** Mainstream economic theory holds that since the exchange rate is fixed there would be *little or no element of speculation* in currencies, since there is little movement in the rates. In reality, most of the currency crises of the post-Bretton Woods period have been the result of speculative attacks on pegged currencies which were considered overvalued. (Refer to previous case study on the British pound in 1992.)

Disadvantages of fixed exchange rate regimes

"…can't win with your hands tied…" ("Fight from the inside", Queen album *News of the World* 1977)

- **Loss of domestic monetary policy freedom:** When a country commits to keeping a certain exchange rate, the central bank will have *limited freedom in setting interest rates* in order to influence the domestic economy. Interest is one of the tools which a central bank can use to keep a peg towards another currency, since higher interest rates attract foreign funds and thus increased demand for the domestic currency. Since the priority of the central bank must be to keep the exchange rate steady there is little room to set interest rates in order to stimulate or deflate the

domestic economy via aggregate demand. There is a *trade-off* between having a fixed exchange rate and being able to set domestic monetary policy – which in extension means a trade-off between exchange rate stability and unemployment.

- **Need of large foreign reserves ('war chest'):** In order to maintain a fixed exchange rate, the central bank will need *ample foreign reserves* for market intervention. Even if this 'war chest' is never used, there must be reserves enough to dissuade speculators from attacking the currency.

- **Possibility of increased unemployment:** If a country runs the risk of continued current account deficits, subsequent downward pressure on the currency will create an incentive to raise interest rates to increase demand for the domestic currency. Government spending might also be cut to decrease imports. Both these policies will have *contractionary effects* on aggregate demand and thus employment.

- **Possibility of 'imported inflation':** When prices rise in trade partner economies relative to the domestic economy, there is a risk that firms and households will simply have to pay more for imported materials and goods, creating *inflation*.

Advantages of floating exchange rate regimes

- **The balance of payments automatically adjusts:** When a country runs a current account deficit, the supply of the domestic currency will increase on the foreign exchange market which in turn will lower the exchange rate. Thus, imports will become more expensive and exports relatively cheaper – the *current account will be self-correcting.*

- **No large foreign reserves necessary:** The central bank will *not need to intervene* by buying/ selling its domestic currency. This means that a large foreign reserve is not needed.

- **Freedom in domestic/monetary policies:** Not having to focus on setting interest rates to achieve a given exchange rate means that the central bank is *free to use monetary policy* to pursue domestic goals of growth and unemployment. This is perhaps the central argument for floating exchange rates.

TRADE

- **Reduced speculation:** Floating currencies are intensely traded by the minute around the world. The availability of information to all traders limits the degree to which one trader will have information that others don't. The past thirty years have shown how floating currencies in fact *do not create speculative crises* – something the famous economist James Tobin noted a few years ago.[7]

- **Less risk of imported inflation:** If trade partners' inflation rates rise relative to domestic inflation, then the domestic economy's goods become relatively cheaper. Hence there will be an increased demand for the domestic currency and an appreciation of the domestic currency. An appreciation of the domestic currency means that *import prices fall* – or at least do not rise as much as in the other countries. Thus there is less risk of importing inflation.

Disadvantages of floating exchange rate regimes

- **Instability and lack of predictability:** Firms often lock international deals far in advance of payment. Not knowing what the actual cost is going to be – due to exchange rate fluctuations – *can deter international trade* and investment. (However, most firms actually set a given exchange rate for future payment, called hedging.)

- **Lack of monetary/fiscal discipline:** Since governments are not forced to keep the exchange rate steady, there is less incentive for governments to keep *budget deficits and inflation* under control. Basically, government can pursue domestic expansionary/ inflationary policies in the knowledge that the depreciation of the currency might serve to automatically adjust any resulting current account deficit in the longer term. However …

- **Loss of competitiveness and efficiency:** … there might be a *loss of competitiveness* over time since the domestic economy increasingly relies on depreciation to remain internationally competitive rather that being forced to innovate and increase productivity.

Preparing for exams

Short answer questions (10 marks each)

1. Using a suitable diagram, explain how a change in a country's imports and exports can affect the exchange rate.

2. Why might a country have difficulty in attaining full employment whilst keeping a current account surplus?

3. Using a diagram, explain how a country can peg (fix) its currency to another currency.

4. Explain why a country's currency might appreciate.

5. Analyse the possible effects of speculation on exchange rates.

6. The central bank in a country raises interest rates. How might this affect this country's currency and balance of payments?

Extended response questions (25 marks each)

1. a. Examine the factors that influence a country's exchange rate. (10 marks)

    b. How might a change in the exchange rate affect the domestic economy of the country? (15 marks)

2. a. Explain the difference between a floating and managed exchange rate. (10 marks)

    b. Discuss the advantages and disadvantages of having a managed exchange rate regime. (15 marks)

TRADE

Summary & revision

1. The type of **fixed exchange rate system** used today is a pegged regime. A country pegs the price of its currency to another currency or to a basket of other currencies.

2. When a currency is readjusted in a band downwards, one speaks of **devaluation**. If the central bank realigns the currency at a higher value, there has been a **revaluation** of the currency.

3. A central bank keeps the home currency pegged to another by **Forex intervention** and changing the **rate of interest**.

   a. *To lower the price* of the home currency to keep the exchange rate within a band, the central bank can sell the home currency (increasing the supply of the home currency) and/or decrease the interest rate (decreasing demand for the home currency).

   b. *To raise the price* of the home currency above the floor set by the central bank, buying the home currency and/or increasing the rate of interest will both cause demand for the home currency to increase.

4. A **managed exchange rate** regime (also known as 'dirty float' or 'managed float') entails periodic government intervention to adjust an otherwise floating exchange rate. There are few – if any – *pure* examples of floating exchange rate regimes.

5. An **'overvalued' currency** can have positive and negative effects.

   a. *Positive effects* include; importers and firms which import costly factors will see increased profit margins; households will have greater spending power; government external debt decreases; there is an incentive for export firms to increase productivity in order to remain competitive internationally.

   b. *Possible negative effects* are; adverse effect on the current account in the balance of payments; higher unemployment in strongly export-orientated economies; repatriated foreign profits are worth less.

6. An **'undervalued'** currency has positive and negative effects:

   a. *Positive effects*; lower price of exports and higher price of imports can improve current account; job creation takes place in export industries; repatriated profits from abroad are worth more at home.

   b. *Negative effects*; import firms see profit margins fall; households' spending power decreases; higher debt servicing of the foreign national debt that is external; possibility of inflation in export oriented countries.

7. There are some notable **advantages and disadvantages of fixed/pegged exchange** rate regimes.

   a. *Advantages* of a pegged exchange rate; creates predictability and certainty for firms' investment plans and households' consumption plans; stability in exchange rates encourages trade; fixed exchange rates mean that governments are forced to exercise fiscal and monetary discipline in order to keep the long run exchange rate; less risk of currency speculation

   b. *Disadvantages*; there is a major trade-off in policy goals since the interest rate must be used to keep the exchange rate fixed and cannot be used to influence AD in the domestic economy; the central bank needs to keep very high levels of currency reserves; large current account deficits put downward pressure on the exchange rate which might cause unemployment when interest rates are raised to countermand this; the possibility of 'importing' inflation when imports rise in price and the exchange rate remains unaltered over time.

TRADE

9. Advantages and disadvantages abound also in floating exchange rate regimes.

   a. *Advantages*; the balance of payments will automatically adjust as large imports will lower the exchange rate and thus induce more exports; no large reserves of foreign currencies necessary in the central bank; freedom in monetary policies as the interest rate can be used for expansionary/contractionary policies; reduced speculation has been the case in most floating regimes; there is less risk of importing inflation as relative inflation rates will adjust the exchange rate accordingly.

   b. *Disadvantages*; instability and lack of predictability for firms and households can deter both trade and investment; governments/central banks might resolve internal unemployment via the 'printing press', e.g. there might be a lack of monetary discipline; countries in floating regimes might rely on falling exchange rates rather than attempting to increase international competitiveness via increased productivity.

TRADE

3.3

70. The Structure of the Balance of Payments

Key concepts:

- The purpose of the balance of payments
- Current account
- Capital account
- Balance of payments equilibrium

"Nothing, however, can be more absurd than this whole doctrine of the balance of trade, upon which, not only these restraints, but almost all the other regulations of commerce are founded."
Adam Smith[1]

Read Smith's quote above. He was reacting to the mercantilist view of the 1600s which basically was that exports are good and imports are bad and that governments should do everything in their power to increase export revenue and decrease import spending. Domestic employment and firms were vociferously protected via tariffs, quotas and many forms of pure prohibition against imports, skilled labour was not allowed to immigrate and countries often limited *immigration* of skilled labour and valuable factors of production...wait, I think I've just described Indonesia!

The mercantilist mindset was focused on achieving a surplus in the balance of trade (here the current account) and this mode of thinking is unfortunately rather prevalent today. It bears stating very clearly at the outset here; exports are nothing more

than *payments* for imports! Stated differently, imports are the revenue for our exports. A nation trying to limit imports is really like a scene out of Monty Python's *Life of Brian* where the gourd seller demands that the buyer haggle down the price! This is a very difficult concept to wrap one's head around and I've actually more difficulties in explaining this to colleagues than to students. Like Keynes said, to learn something new one must often first un-learn the old.

The purpose of the balance of payments

The balance of payments is a record of a country's money in- and outflows to and from the foreign sector during a period of time. It is basically a *ledger* (= book for keeping financial records) in every country where one records all the inflows and outflows of money from and to the foreign sector. Inflows can consist of export revenues while corresponding outflows would consist of payments for imports. Other in- and outflows could be investment flows from/to other countries. In essence, the balance of payments accounts for a country's income flows, investment flows and asset base in dealing with the foreign sector. The balance of payments categorizes these flows into three components; the **current account**, the **capital account** and the **financial account**. An inflow to the balance of payments is termed a *credit* item and an outflow a *debit* item.

TRADE

---

1   Adam Smith in *Wealth of Nations*; quoted from *Irwin*, page 90

497

Another childish example coming up now. If you have an income (from exports) of £100 during a given time period and expenditure of £90 (on imports), then by the end of the period you have increased your *net* holdings, i.e. *your net assets have increased*. Assuming that your export earnings go towards £90 worth of imports and you hold your £10 in another country, your net *foreign* assets have increased. This would be a *net inflow into current account* (credit items > debit items) and a *net outflow from the financial account* (debit items > credit items). If, however, you spend £110 on income of £100 then the £10 which did not flow in during the time period must come from somewhere; loans from a friend or the selling of your economics textbook. In any case, the 'overspending' means that your *net assets have decreased*. If you borrowed from abroad, your net foreign assets have decreased. You would have a current account deficit and a financial account surplus.

The 'purpose' of the balance of payments is to account for all money flows to and from the foreign sector arising from the trade of goods and services, investments and income thereof, and other currency flows arising due to loans, deposits and currency speculation. It pinpoints a nation's competitive strengths or weaknesses in the international sector and is an indicator of both a nation's economic solidity in the long run and of the underlying strength of the currency. Basically, a country running a solid current account surplus must have increased its net foreign assets and vice versa if there is a current account deficit. The balance of payments also pinpoints where the money is coming *from*; where it is *going to*; and *why*.

Current account[2]

The current account is a country's record of all money flows arising from economic activity with foreign countries; exports and imports of physical goods plus exports/imports of services and other incomes and transfers.

Moving on to a bit more depth, Figure 70.1 gives the balance of payments for Ireland in 2011. The current account has been sub-divided into two main components:

The **visible trade balance** [3] comprised of merchandise exports [1] and imports [2]; and the **invisible trade balance** [13] which in turn is comprised of net flows in services [6], income [9] and transfers [12]. Please note that the balance of payments is an 'end-of-game snapshot picture' of total flows which have occurred *during a time period*, e.g. over the course of a year.
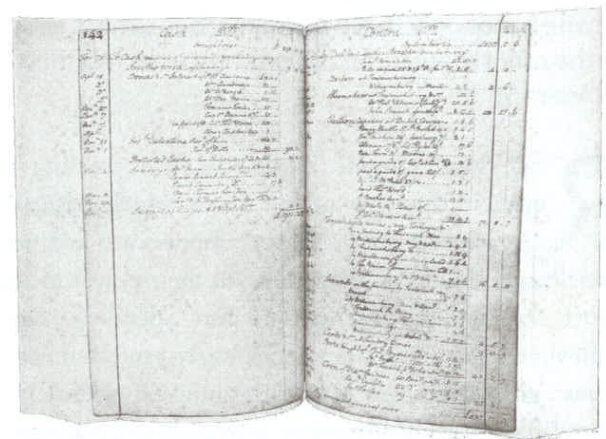
George Washington's Cash Book

---

2   I once again have a debit in my personal current account to the Central Statistics Office of Ireland (*An Phríomh-Oifig Staidrimh* in Irish). The balance of payments figures for national income vary enormously from country to country and can be immensely difficult to plough through. I owe Mary Bourke a few pints for taking the time to help a confused expatriate Swede with the numbers.

TRADE

Figure 70.1 Ireland's balance of payments for 2011

Current, Capital and Financial Account 2011 (millions EUROs)

	Current account	€m		Capital account	€m
1	Exports of goods	85,258	15	Net capital transactions (transfers plus acquisition of non-produced/non-financial assets)	−429
2	Imports of goods	48,862			
				Financial account	
3	Visible trade balance	36,397	16	Net FDI flows	10,968
			17	Net portfolio investment	28,186
4	Exports of services	79,084	18	Net Other financial flows	−32,108
5	Imports of services	82,291	19	Reserve assets	341
6	Balance of trade in services	−3,207	20	Financial Account balance	7,389
7	Income receipts from abroad	56,433			
8	Income payments abroad	88,597			
			21	Balancing item/stat. discrepancy (Net errors and omissions)	−7,087
9	Net income from abroad	−32,163			
10	Current transfers from abroad	4,936			
11	Current transfers paid abroad	5,838			
12	Net transfers from abroad	−902			
13	Balance on invisibles (items 6, 9, 12)	−36,272			
14	Current account balance	127	22	Capital and financial account balance (items 15 and 20 minus 21)	−127

Visible trade balance

When an Irish good is sold to a foreigner, goods flow out and money flows in, while goods from abroad sold in Ireland create money outflows. One calls an inflow a **credit** in current account and an outflow a **debit**. (I will use credit and debit initially, so that you recognise the terms in other literature, but will thereafter confine myself to the use of 'inflow' and 'outflow'.)

Exports: When my Irish friend Joe Collins in Tanzania gets homesick for Ireland and his wife goes down to the local liquor store in Dar es Salaam for a bottle of Irish whisky, the effect is to **credit** (+) the Irish current account in merchandise exports [1]. When the Collins family goes home to Ireland and visits the family pub, there will also be a credit in Irish exports since Joe lives and earns money in Tanzania not in Ireland.[3]

Imports: When Joe's father, Michael, buys foreign beers for the pub in Ireland, there is a **debit** (-) in the Irish current account in merchandise imports [2].

3   In actual fact, Joe gets a rather good price on food and drink there.

Adding up all the other millions of visible trade transactions with the foreign sector during 2011, i.e. the sum total of exports [1] minus the sum total of imports [2] we get a **visible trade balance** [3] of €36,397 million. This is a (visible) *trade surplus*. Basically, the Irish have earned more from exports of goods than they have spent on imported goods.

> **Definition: 'Trade balance'**
> The sum of visible export revenue minus the sum of visible import expenditure equals the (visible) trade balance. A negative value is called a trade deficit while a positive value is a trade surplus.

Invisible trade balance

The same principles as in the examples in merchandise trade above hold true for trade in *services* and other invisible (i.e. intangible) goods. The **invisible trade balance** is comprised of a variety of flows arising when money passes a border for reasons other than for the payment of merchandise.

- **Tourism/services**: Tourist services are exported whenever they are being bought by foreign money, for example when Uwe from Germany stays at O'Leary's Bed and Breakfast in Adare or when Hiroshi from Japan takes a boat trip in Dingle Bay to see Fungi the dolphin. Other services are when Ireland's Allied Irish Banks (AIB) sells financial services in New York, or when Ryan Air sells flights to French people – both will entail service export income for Ireland. Conversely, when Irish people buy foreign services there will be an outward flow of income – an import of services – in the Irish balance of payments. The net sum of these flows is a deficit in the balance of services [6] of -€3,207 billion.

- **Flow of incomes generated abroad**: Irish economic activity abroad will generate foreign earnings. When these are *repatriated* (= brought home to the mother country)[4] the Irish current account is credited.

  - When Irish firms make a **profit** abroad and repatriate it, there is an inflow of money. A foreign firm in Ireland repatriating profits is of course an outflow in current account.

  - This is also the case when Irish savings abroad generate incomes which are repatriated, such as **dividends** from foreign shares and **interest payments** on foreign bank accounts and bills/bonds. Naturally, there will be outflows from Ireland as foreigners repatriate incomes generated by their ownership of Irish capital.

The net sum of these outflows is shown by the **net income paid abroad** [9] of - €32,163 million.

- **Net transfers**: Ireland will have a number of additional money flows which are recorded in the current account.

  - For example, if my friend Joe in Tanzania sends some of his income home to his parents every month, this **remittance** is shown as an inflow in the invisibles section of the Irish current account. Basically, Ireland has exported labour in the form of an IB math/physics teacher. (See *Applied economics* below.)

  - Other inflows to Ireland are **farm subsidies** from the EU and **project grants** from various international organisations.

  - Naturally Ireland will also have an outflow of transfers: **government dues** to the EU and UN; **grants to Irish students** overseas; **aid monies** paid directly and indirectly to recipient countries ... etc.

The **net outflow** of these transfers [12] was - €902 million in 2011.

> **Definition: 'Invisible trade balance'**
> The net sum of the trade in invisibles, e.g. services, incomes and transfers, equals the **invisible trade balance**.

Current account balance

The sum of the visible trade balance [3] of €36,397 plus the invisible trade balance [13] of -€36,272 shows the net in- and out-flows of money to Ireland in current account during 2011. Ireland shows a net inflow (credit) of €127 million during the year, in other words a **current account surplus**.

---

4   I've always wondered if '...*repatriate* to the *mother* country ...' is a contradiction in terms.

Applied economics: 'Migratory cash'

Total remittances in the world in 2006 were close to USD300 billion – an 83% increase over five years where some USD240 billion of the total goes to developing countries. For developing countries this represents close to one third of all financial flows. Figure 70.2 below shows the top ten recipients and sources of workers' remittances from abroad during 2006. It is clear that monies sent home by citizens working abroad can have a colossal effect on the current account in the balance of payments. For example, the $US27 billion going to India represents over 3% of India's GDP and 37% of total net invisibles in India's current account. In the US, the world's largest importer of foreign labour, the $US42.2 billion leaving the country as remittances constitutes close to 5% of the *total* USD856 billion current account deficit of 2006.[5]
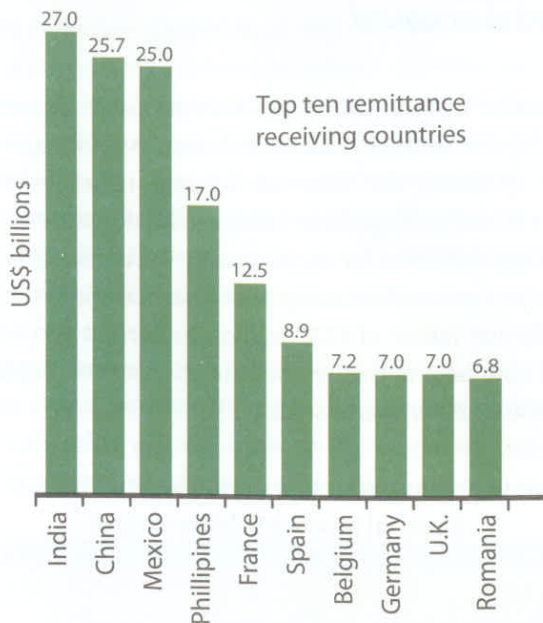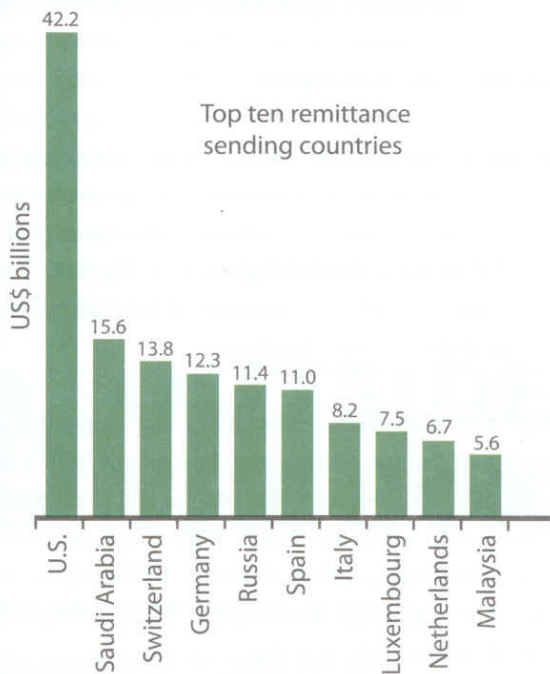
Figure 70.2 Top ten foreign remittances in selected countries (2006)

(Source; World Bank, 2008 Remittance Handbook, Top Ten)

Capital account[6]

The 'other side' of the balance of payments shows what Ireland has done with the €127 million in the current account surplus. The capital account is a rather small part of the balance of payments (about 0.6% of overall net flows in this example) and consists of two parts:

- **Capital transfers**; primarily government and non-government debt forgiveness and the transfer of assets due to immigration and emigration.

- **Acquisition or disposal of non-produced and non-financial assets**; transfers of funds for the purchasing or selling of land and the transfer of funds relating to royalties, patents and other forms of intellectual property rights.

TRADE

---

5   International Organization for Migration at http://www.iom.int/jahia/Jahia/about-migration; Migration News at http://migration.ucdavis.edu; World Bank at http://siteresources.worldbank.org/INTPROSPECTS/Resources/334934-1199807908806/Top10.pdf; Bureau of Economic Analysis (BEA) at http://www.bea.gov/newsreleases/international/transactions/transnewsrelease.htm

6   For details on this rather tricky subject, see the IMF standardisation at http://www.imf.org/external/np/sta/bop/BOPman.pdf

Financial account

The *financial account* shows all other forms of money flows not arising due to trade, income and returns on foreign assets. Broadly speaking, the financial account tallies (= counts) investment and savings flows between the domestic economy and foreign countries. In the case with Ireland in 2011, there needs to be a net outflow in capital + financial account to make up for the net inflow of €127 million in current account. The cardinal rule here is that the sum of the current, capital and financial account is always zero!

Current account balance + capital account balance + financial account balance = zero!

Definition: 'Capital and financial accounts'

The capital and financial accounts are the mirror image of the current account, showing how capital transfers flow to and from other countries. The capital account shows the change in net foreign assets of a country due to foreign direct investment (FDI), portfolio investment (shares, bills, bonds), currency flows, loans, and deposits. A current account surplus means that there must be a net outflow in capital + financial account, e.g. an increase in net foreign assets.

In keeping with my rather sad habit of using childishness to get the message across, consider that an account of your income, spending and asset flows during a month are as follows:

Income:

£90 (+)   Bought 65,000 shares in AIG: £5 (−)

Spending:

£80 (−)   Loan to a friend: £5 (−)

'Under-spending':

+£10   Net increase in assets: −£10

Using less childish economic terminology now; as your 'export earnings' (income in the figure above) are larger than your 'import expenditure' (spending above) by £10, you have a *current account surplus*. You used this money to buy shares and lend money to a friend. As you now have £5 worth of shares and a £5 claim on someone else's assets, your *net* assets

have increased by £10, which is a *capital + financial account deficit*. The trick here is that you must readjust your thinking a bit when looking at the minus and plus signs in the capital/financial accounts:

- A plus sign (credit or inflow) in the capital/financial account means that money has flowed into your economy and that there has been a *decrease in net foreign assets*.

- A minus sign (debit or outflow) in the capital/financial account therefore means that *net foreign assets have increased*.

Recall that Ireland had a current account surplus of €127 million during 2011, i.e. a net inflow. The capital and financial accounts show where that 'additional' €127 million went to – in effect, which assets the Irish are buying abroad. It is important to realise that all the values given in the capital/financial accounts are **net values** and that total in- and out-flows over the course of 2011 were hundreds of times larger.

Net foreign direct investment (FDI): Investment abroad is either direct investment or indirect investment.[7] An Irish company building a plant or buying an existing factory in Germany entails *direct* investment. The positive value of €10,968 [16] here shows that more direct investment money is flowing in to Ireland than abroad.

Net portfolio investment: If an Irish company/resident buys German shares, bills or bonds, one speaks of *portfolio* investment abroad; if a German buys Irish shares there is an inflow (credit) in the Irish financial account. The €28,186 billion net inflow from Ireland [17] means that Irish firms/residents have increased their foreign holdings.

Net other financial flows: The Irish will deposit some of their savings in foreign banks; Irish banks will lend money to foreigners; and Irish financial institutions will place short term monies abroad in order to earn a higher rate of interest, or simply speculate that a certain currency will increase in value. These are short term investments and the €7,389 million in net financial inflows [18] signifies a *net decrease* in short term foreign assets for the Irish – which is the same as an increase in short term liabilities to foreigners. In other words, the net

7 The rule for Ireland being that if shares in a firm account for more than 10% of the total value of shares, it is direct investment rather than indirect – portfolio – investment. See details at the Irish Central Statistics Office: http://www.cso.ie/en/media/csoie/releasespublications/documents/economy/2011/bop_q42011.pdf

TRADE

balance in financial inflows of €+7,389 billion shows an *increase in foreigners' short term holdings* in Ireland.

Foreign reserves: All countries' central banks hold an amount of foreign currencies and, traditionally, gold. The positive value of €341 million [19] means – rather confusingly – that reserves at the Bank of Ireland have *fallen* by this amount. Somewhat simplified, foreign reserves have been used to buy up EUROs, shown in the balance of payments as an inflow (credit) in net foreign reserves. In any case, using a positive value in the financial account to show a *decrease in foreign reserves* is in keeping with standard bookkeeping procedures; a net inflow in capital/financial accounts means that domestic holdings of foreign assets have fallen.

Balancing item: The millions of daily money transactions between Ireland and other countries is an almost impossible task to precisely account for. There will be statistical errors, time lags in accounting, and a number of money flows which simply will not show up due to tax/tariff avoidance, parallel markets and illegal trade. Since the balance of payments must always balance, a **balancing item** or **statistical discrepancy** ('the bit missing') of -€7,389 million [21] has been subtracted.

Balance of payments equilibrium

The Irish current account surplus of €127 means that Irish households and firms have increased their net foreign holdings by the same amount; the sum of the capital and financial accounts shows a net outflow of -€127 to the foreign sector. I agree, this is hideously boring … so I am going to finish of with a few silly examples. *You* might not enjoy them but it keeps me going at the keyboard.

OUTSIDE THE BOX

Humans are from Venus, economists from Mars

A question that all students of balance of payment accounts ultimately ask themselves is of course "Do we trade with Mars"? This is a good question, posed by several high-ranking economists.[8] Here's why the question arises…

---

8   See for example Abel and Bernanke, page 177

Assume only two countries, A and B, who trade with each other; one country's outflows on balance of payments would be the other country's inflows. A's exports would be B's imports; B's FDI would flow out of the financial account and into A's financial account … etc. What would the *sum* of both countries current account be?! Yes, zero. Adding 190 more countries do not change this; the sum of the world's current accounts (or capital + financial accounts) must be zero.

It turns out, however, that the world has been running a *current account deficit* for over 30 years, and since we in fact do not trade with other planets, there must be another explanation. Think before you check the answer in the footnote.[9]

I finish this chapter by giving a few schematic examples of balance of payments flows between two countries. Let's look at the Irish and Australian economies which are now happily trading goods and service instead of prisoners and colonial taxes.[10] Assuming that the two countries only trade with each other, here is a selection of economic transactions taking place between the two economies during a few months (using country-neutral 'dollars'):

[1]   A distiller in Ballyragget, Ireland, sells $2,500 worth of Top 'a The Mornin' Whiskey to a pub in Cootamundra, Australia.

[2]   Goanna-Mick from Meekatharra, Australia, on vacation in County Mayo, Ireland, rents a cottage overlooking lots of sheep in Ballyhaunis for a week at $2,000.

[3]   Mary Bourke from Ballydehob, Ireland, tires of all the silly requests she gets working at the Central Statistics Office and puts $6,000 into starting a subsidiary legal firm in Sydney, Australia.

---

9   The most likely explanation is that a portion of interest, profits and dividends earned abroad simply are not reported at the receiving end (inflow into current account) in order to evade taxes. However, the funds will frequently be registered in the country sending the money (outflow from current account). This systematically *understates* the current account inflows in many countries, leading to a 'global current account deficit.'

10   Thousands of Irish 'troublemakers' and convicts were evicted from Ireland and sent as forced colonisers to Australia in the 19th century. This popular export commodity has resulted in tens of thousands of present-day Irish descendants in Australia with names like my Australian editor; 'McAuliffe' (which in fact reads, 'son of Olaf', which means he is actually a Scandinavian descendant!).

[4] Bruce from Kingoonya, Australia, is prolonging adolescence at university in Galway (just west of Ballinasloe), Ireland. He receives a $3,000 student grant from an unwitting municipal council back home.

[5] Sean from Ballyroan, Ireland, buys shares in the Areyonga Shipyard for $3,500. (Sean really needs to get a good map of Australia.)

[6] Alice from Three Springs, Australia, puts $6,500 into a bank account in Ballycanew, Ireland.

[7] Mrs Bourke in Ballydehob, Ireland (from [3] above), repatriates $500 in profits from her first customer (an Irish descendant of the McGee Clan out on bail) in Sydney.

[8] Bush Tucker International Delicatessen in Bulgroo, Australia, sells $5,000 worth of sliced baby koala and pickled platypus nibbles to the local Greenpeace chapter in Ballyconnell, Ireland.

These international flows of money between the two countries are shown in Figure 70.3 where I have simply bundled the capital and financial accounts together as the distinction is not really important here. Following the flows, it is clear that any outflow (debit) of money from Ireland leads to a corresponding inflow (credit) in Australia, and vice versa. Irish exports of Whiskey and tourist services are inflows to the visible and invisible trade balances in the Irish current account – which are of course outflows in the Australian trade balances in the Australian current account. Irish investment in Australia gives outflows of money in the Irish financial account and concurrent inflows to the Australian financial account.
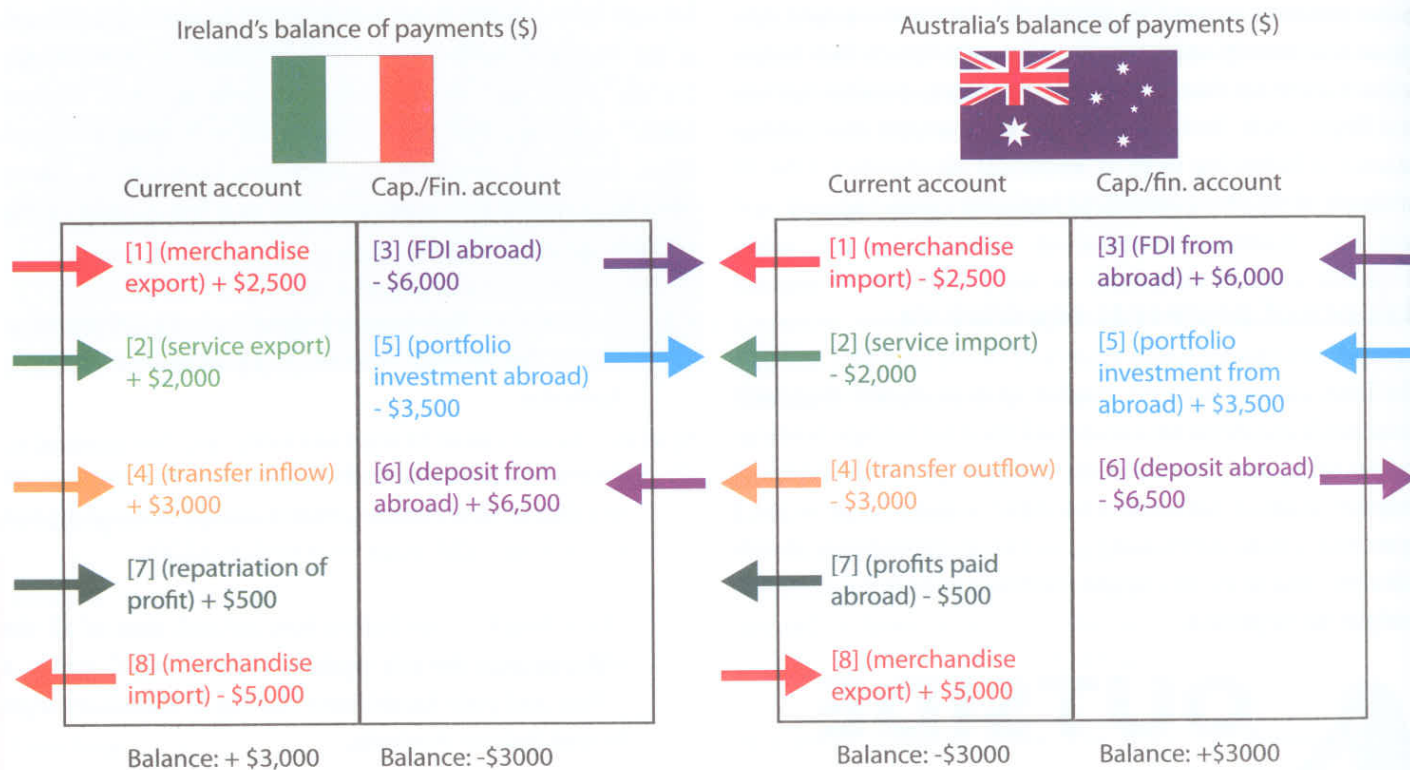
Figure 70.3 Make-believe balance of payments for Ireland and Australia

Ireland has a current account surplus of $3,000 and thus a corresponding capital/financial account defici. Once again, do not be confused by the negative value in the Irish capital/financial account! An outflow here means that Ireland has increased its net foreign assets; Ireland has increased holdings in Australia more than Australians have increased holdings in Ireland. Australia has a current account deficit, concomitantly. The $3,000 worth of imports in excess of export revenue has apparently been funded by an inflow of investment money from Ireland.

Is a current account deficit bad for a country?

This is a good question and one which is vociferously (= loudly) debated amongst economists and in the media. In simple terms the answer is "no, not always". A more balanced answer is that since a current account deficit means a capital/financial account surplus, it depends on *why* money is coming in on capital/financial account and *what* the money is being used for. In the story above, Australia's current account deficit has evidently been largely financed by the willingness of Ireland to invest in Australia. Looking at real life examples (and done in Chapter 72) the Australian economy has benefited enormously from the foreign investment inflows; growth rates have been high and unemployment rates low throughout the financial crisis. Generally speaking, as long as a current account deficit is made up by *investment* inflows on financial account there is little cause for worry in the short to medium term.

It is when *loans* and *speculative monies* flow in on capital account to fuel *consumption* that harmful effects will show, as the country would in effect be living beyond its means and would someday have to pay back loans.[11] There are also long run considerations. For example, if foreign companies start pulling out of a country – say, US firms start shutting down in Ireland due to severe recession in the countries where the FDI is coming from – then there could be serious repercussions for the Irish economy. This is *exactly* what happened as we shall see in Chapter 72.

Summary & revision

1. The **balance of payments** is a record of all in- and out-flows of money in an economy during a period of time (one year) arising due to economic transactions and financial dealings with other economies.

2. The balance of payments consists of the **current, capital** and **financial** account

3. The balance of payments **must always balance**, e.g. current account = capital account + financial account. The sum of the three accounts must equal zero.

4. **Current account** is commonly divided into a *visible trade balance* (export revenue minus import expenditure) and an *invisible trade balance* (net balances of trade in services, net income from abroad and net transfers from abroad).

5. **Capital account** consists of *capital transfers* (such as government debt forgiveness and assets transferred by immigration/emigration) and acquisition of *non-produced and non-financial assets* (land ownership and intellectual property rights).

6. The **financial account** shows the change in net foreign assets/liabilities. It consists of the balances of *FDI* (net change in physical capital such as factories and shops), *portfolio* investment (shares, bills and bonds), *financial flows* (such as currency speculation and bank deposits) and the *foreign reserves* (foreign currencies, gold and deposits with the IMF held by the central bank).

7. The **balancing item** (or statistical discrepancy) is a post that is inserted to ensure that the accounts balance. It is a measure of accounting uncertainty and error.

8. A **negative balance on capital and financial account** means that there is a current account surplus – and vice versa.

TRADE

---

11 It's a bit like a household borrowing money. If the money goes to rebuilding the porch and repairing the roof, then the house will be worth more in the future, e.g. there is a return on the spending. Hence, there is no harm in borrowing money. If, however, the borrowed money goes to pay for Christmas presents … go figure.

71. Current Account and the Exchange Rate

Key concepts:

- Current account deficit and the exchange rate
- Current account surplus and the exchange rate

We have gone through how the exchange rate can affect the current account; a depreciation lowers export prices and increases import prices. This is a two-way street however as deficits or surpluses in current accounts will have an effect on the exchange rate. Underlying or causing the current account deficit/surplus one often finds increased national income – both from trade partners or domestically. In Chapter 67 the link was made between a) rising domestic GDP and increased imports and b) rising GDP in trade partners' economies and rising domestic exports. Throughout this chapter I will use the US economy as an example. (Remember that the US dollar is in a *floating* exchange rate regime.)

Current account deficit and the exchange rate

When American exports decrease, there will be a decreased (derived) demand for the US dollar as importers in other countries will not need as many dollars to buy American goods. The same goes for services and tourism; less demand for US banking services and tourism in the US mean less demand for the US dollar. As for a US current account deficit mainly caused by US imports, increased imports mean an increase in the supply of the US dollar and likewise a depreciation of the US dollar.

- **Decreased US exports** will decrease the demand for the US dollar, and subsequently cause an appreciation of the dollar.

  - ○ Δ↓US exports → Δ↓ demand for USD → depreciation of the USD

- An **increase in imports** and/or an increase in American tourism abroad means that more dollars will be traded to purchase imports and buy tourist services. This

increases the supply of dollars on the market and the US dollar depreciates.

  - ○ Δ↑US imports → Δ↑supply of USD → depreciation of the USD

Current account surplus and the exchange rate

You know the drill now so this will be given short shrift:

- **Increased US exports** will increase the demand for the US dollar, and subsequently cause an appreciation of the dollar.

  - ○ Δ↑US exports → Δ↑ demand for USD → appreciation of the USD

- A **decrease in imports** means that fewer US dollars will be traded (per day or week) to purchase imports and go on holiday abroad. This **decreases the supply of dollars** on the market and the US dollar appreciates.

  - ○ Δ↓US imports → Δ↓supply of USD → appreciation of the USD

In summa; a *current account deficit would be associated with a weak US dollar* and a current account surplus linked to a strong US dollar. Over time this would help the US balance of payments attain equilibrium:

- A weak US dollar would ultimately decrease imports and increase exports … moving the current account towards balance.

TRADE

- A strong US dollar would have a negative effect on exports and increase imports ... again, the current account moves toward equilibrium.

This argument is one of the arguments in favour of floating exchange rates, namely that in a floating exchange rate regime there is a built-in long run force moving the balance of payments towards equilibrium. See Chapter 69.

Note how difficult it can be to see cause and effect! The above theoretical discourse proposes that increased exports (or decreased imports) will *cause* an appreciation of the US dollar (positive correlation between current account balances and the exchange rate). Causality in the opposite direction is of course quite possible – an appreciation of the US dollar would *cause* a worsening of the US current account (negative correlation between current account and the exchange rate). Whatever the outcome, we expect correlation between the two variables and this leads us to ...

... One big massive *however* ...

It looks good in theory but in using that most reliable of tools, real life data, how strongly correlated is the US current account to the exchange rate for the US dollar?

Quite frankly; it isn't. Figure 71.1 plots out the US dollar exchange rate (trade weighted index, right axis) and the US current account (as a percentage of GDP, left axis). There is in fact very little *consistent* correlation over the 25 year period. The period 1980 to 85 shows negative correlation, then positive up to 1988, negative to 1990 ... etc. My main point here is that there is a) no *simultaneous* correlation and b) when there is correlation it is mostly negative rather than positive – which means that the link between current account surpluses and appreciation is very weak.[1]

Here is an experiment. Grab your ruler again and slide it along Figure 71.1. Better still, get a piece of overhead plastic and trace the US dollar index curve. Now you can shift it back and forth to see whether there is a lag between the change in the exchange rate and the current account.

Try moving the overhead plastic 'four years' to the left ... comments?[2]
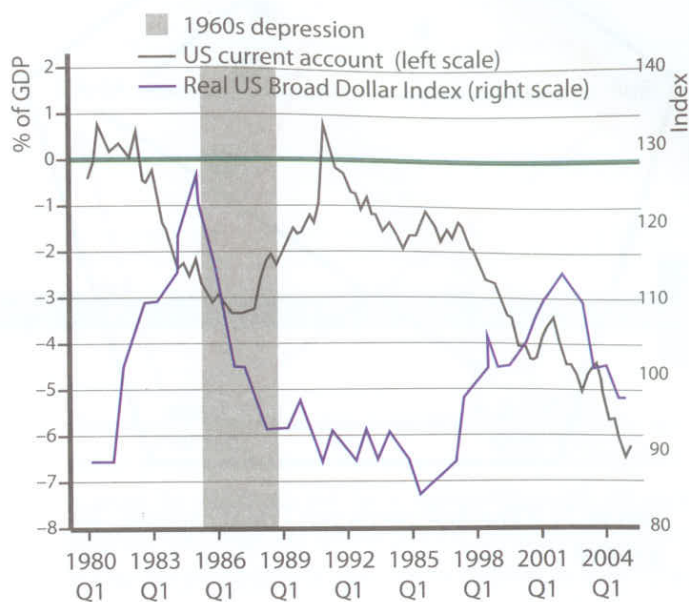
Figure 71.1

Exam help;

The 'Balance of payments Pentagon'

As shall be shown in the remaining Section 4 chapters, the balance of payments will have an effect on other macro variables, primarily exchange rates, aggregate demand and the terms of trade. But it doesn't stop there – said macro variables will in turn influence the balance of payments! And finally, the macro variables will influence each other. It's a big can of worms, but I shall attempt to give you a few examples. Figure 71.2 below illustrates the 'multi-causal' nature of macro variables that are linked to the balance of payments (BoP in the figure).

TRADE

1 There is, however, rather strong correlation between the financial account in balance of payments and the US dollar exchange rate! This is not surprising since the US attracts huge amounts of inward investment and has numerous subsidiary companies abroad. See http://www.gold-eagle.com/editorials_05/bart070106.html

2 There is improved correlation and an indication of causality; with a four year lag, there is evidence in the 1980s and 2000s of positive correlation. The lag indeed indicates that the exchange rate is causally linked to the current account balance – a change in current account *causes* a change in the exchange rate after a three to four year lag.
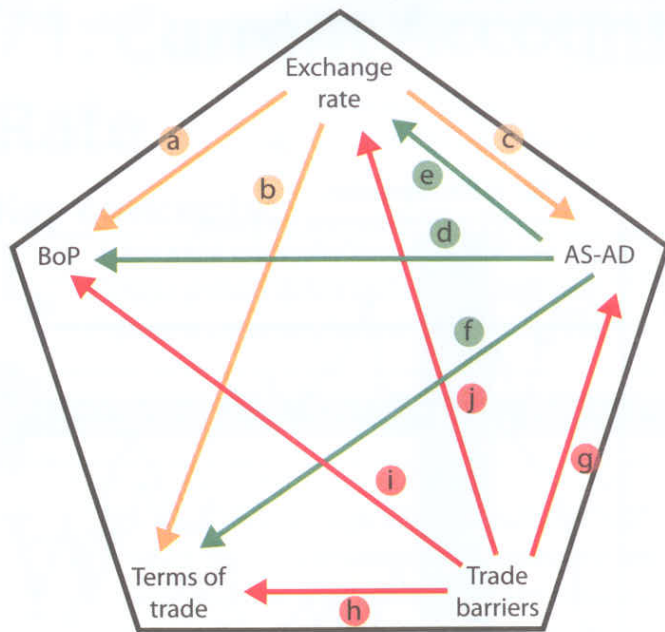
Figure 71.2 The BoP pentagon - dual causality all around

Here are a few examples of some cause and effects linked to the balance of payments:

1. A depreciation of the home currency would make exports cheaper (rise in exports) and imports dearer (fall in imports). This would improve the balance of trade in current account (a). The depreciation lowers the price of exports relative to the price of imports, the terms of trade worsen (b). Increased exports increases AD (c).

2. A decrease in demand for exports causes AD to fall and current account to worsen (d) and demand for the home currency to decrease (e) which worsens the terms of trade (f).

3. Increased tariffs lower imports, increasing AD (g), the terms of trade worsen (h), current account is positively affected (i) and increased demand for exports appreciates the exchange rate (j).

This is not rocket surgery: any change in one variable will, well, invariably have an effect on the others. Falling incomes due to contractionary fiscal policies will reduce inflation … cause imports to fall…cause appreciation of the home currency … etc. Ask your math teacher how many possible permutations there are in the pentagon above. When he/she has worked it out, tell them, oh, there are three parts to the BoP … and there is AS also … AD in trade partners' economies …

Occultic Pentagram

Summary & revision

1. A **deficit in current account** is theoretically associated with a *weakening of the exchange rate*:

    a. Decreased exports would lower the demand for the Home currency and lead to a depreciation.

    b. Increased imports would increase the supply of the home currency and lead to depreciation.

2. A **surplus in current account** is theoretically linked to a *strengthening of the exchange rate*:

    a. Increased exports lead to an increase in demand for the home currency and therefore an appreciation.

    b. Decreased imports can cause the supply of the home currency to decrease and cause an appreciation.

3. One of the key economic arguments in favour of a floating exchange rate is that the balance of payments **automatically adjusts** in the long run. A current account deficit over time will depreciate the domestic currency and maker exports cheaper and imports dearer. This would thus help to re-balance the balance of payments via a current account improvement.

TRADE

72. HL extension – Implications of a Current Account Deficit/Surplus

Key concepts: HL extensions

- Calculating current account balance
- Effects of a persistent deficit in current account
- Correcting a current account deficit
- Effects of a long run current account surplus

Calculating current account balance

In Figure 72.1 you will find a simplified current account. Do the standard-issue fill-in-the-blanks and address the questions further on. Note that I forego the 'balancing' version where the current and capital/financial accounts are side-by-side since the version given here is in all likelihood the one you will see in your exams. Answers are in the end of Chapter Summary.

Figure 72.1 Fictitious current account

Current account	€m
Exports of goods	x
Imports of goods	480
Visible trade balance	−20
Exports of services	x
Imports of services	260
Balance of trade in services	80
Income receipts from abroad	160
Income payments abroad	x
Net income from abroad	−60
Current transfers from abroad	150
Current transfers paid abroad	100
Net transfers from abroad	x
Current account balance	x
Capital account	€m
Net capital transactions	5
Financial account	€m
Net FDI flows	x
Net portfolio investment	−200
Net Other Investment	30
Reserve assets	20
Financial Account balance	85
Balancing item/stat. Discrepancy	x
Capital and financial account balance	x

1. How does it seem that this economy is funding the trade deficit primarily? (Since there is a current account deficit, it must be made up for by a capital/financial account surplus. The large net inflow in FDI explains this. There is also the possibility of speculative inflows in 'Net other investment'.)

2. GDP for this economy is €3,400 million. Is the current account situation alarming? (No, it's just under 1.5% of GDP which is historically low.)

3. Would you think that GNP might be larger than GDP for this economy? (Highly unlikely since there is a net income outflow (debit) in current account.)

4. Assume that this economy has a floating exchange rate. What is the possible effect on the exchange rate? (If exports continue to exceed imports, it is likely that the supply of the Home currency is increasing. This can cause a depreciation of the exchange rate.)

5. Assume instead that the exchange rate is *pegged* to another currency. What actions might the central bank be taking right about now? (A current account deficit would put downward pressure on the exchange rate. In order for the central bank to retain the peg towards the foreign currency it might raise interest rates and/or intervene on the Forex market by support-purchasing its own currency.)

6. Does this economy seem strongly export orientated? (I would say so since almost 15% of GDP comes from export revenue. For comparison I point out the corresponding figure for Japan has been between 13 and 16% for over 20 years.)

7. Have the economy's foreign reserves increased or decreased? (Decreased since there is a net inflow in reserve assets which means a *decrease* in net foreign assets, e.g. in the foreign reserves held by the central bank.)

Effects of a persistent deficit in current account

"If something cannot go on forever, it will stop."
Herb Stein, Chair, Council of Economic Advisers, 1972–74

A country running a current account deficit will of course have a corresponding capital/financial account surplus – the country is in effect selling domestic assets to the foreign sector. However, it is far too simple to label a current account *deficit* as 'bad' or harmful and a current account *surplus* 'good' or beneficial; it depends on *why* there is an imbalance.

Exchange rate effects

Perhaps the most immediate and visible effect of a current account deficit is the effect it has on the **exchange rate**. As explained in Chapter 71 a current account deficit means larger import expenditure than export revenue, there will be downward pressure on the exchange rate.

Foreign debt issues

Borrowing from abroad means that the loans will ultimately have to be repaid with interest. Continuous current account deficits will be looked upon harshly by the international business and financial community, and ultimately the **ability to pay off foreign debts** might be questioned. Ratings agencies such as Moody's and Standard and Poor are quick to downgrade bad debtors! Wary investors might choose to avoid a weak economy and this will result in less demand for the home currency as fewer investors/loan-givers are willing to risk putting assets into the country, putting downward pressure on the currency.

Interest rate issues

In the event that the home currency indeed depreciates, the debtor nation could suffer a severe shock as *debt servicing* becomes more costly. The home country will have to *offer even higher interest* to foreign lenders – which together with a depreciated currency can make future debt payments a serious issue. It also puts the domestic economy somewhat at the mercy

of international business cycles and interest rates – something a good many developing countries learned the hard way during the 1970s and '80s.

The domestic economy might in essence be forced into **raising interest rates** in order to attract continued foreign (portfolio) investment and keep a desired exchange rate. In essence, the domestic economy is letting foreign firms fund domestic investment. Ultimately, the sales of domestic assets to foreigners will cause outflows in the form of repatriated profits and dividends – which will in fact intensify the current account deficit.

Demand side issues

If an economy is financing a current account deficit primarily by way of attracting foreign deposits from abroad, then it has to offer an attractive rate of interest. As explained in the previous two points, a debtor nation will in the long run probably have to offer ever higher interest to foreign speculators/investors – a 'risk premium' if you will. Higher interest rates will have a contractionary effect on aggregate demand and the contingent macro objectives of growth and high employment.

Speculative bubbles

The incoming funds on capital account might be due to **speculative inflows** – in which case the recipient country could be in for some serious trouble when these inflows cease. (See *Case Study on South East Asian Crisis* following.) If foreign investors/speculators start to fear that the often highly *speculative* assets they are buying (such as property, shares and currency) in the host country are overvalued, they will leave. As foreign capital starts to exit and FDI and portfolio investment seek other markets, then the recipient country can find itself in a very serious situation where stock markets and/or property markets crash, aggregate demand plummets and domestic unemployment rises drastically. There will, however, also be a significant reduction in imports ... but saying this is a 'positive' is a bit like saying 'Well, cancer cured my smoking!'

The Irish collapse in 2008/'09

The result of investment inflows that suddenly dry up is something that the Irish became painfully aware of during the credit crisis of 2008/'09. Ireland has long attracted a great deal of foreign direct investment (FDI) over the past 20 years by granting firms tax holidays – initial periods during which foreign firms pay no taxes on profits – of up to 10 years, and ultimately a low 12.5% corporate tax rate. In addition to this,

Ireland has strong historical/cultural ties with the US and an abundant, young, well-educated English-speaking workforce. This attracted FDI in massive amounts and US multinational companies (MNCs) accounted for 16% of Ireland's GDP by 1995.[1] By 1998, Irish GDP at PPP[2] had gone from less than 65% of the EU average to over 100% of the EU average.[3]

The over 1,100 foreign companies operating in Ireland generated a great deal of profit which was *repatriated* abroad rather than re-invested in Ireland. It also provided one of the main *tax bases* for the Irish government – income tax was a low 26% of government revenue by 2005. Ireland was also part of the EURO-zone and therefore enjoyed low interest rates set by the European Central Bank (see Chapter 74). Unemployment was around 4 – 5% and the economy grew by almost 70% between 1996 and 2001 – hitting over 10% per year towards the end of the 1990s and never falling below 4% per year between 1995 and 2007.[4]

So; high growth, low taxes, massive inward flows of FDI, low unemployment, low interest rates – and we add in government subsidies for building and you get a *building boom*! By the peak of the boom period in 2006/07, some 23% of Irish GDP was composed of construction. Add in housing speculation and high inflation and you get a highly overheated economy and huge amounts of household debts – people were buying houses to 'flip' them a few months later for a profit. Then the recession hit US and spread to Ireland. FDI and other investment flows to Ireland plummeted, growth turned negative, unemployment skyrocketed and government debt skyrocketed since such a large proportion of tax revenues came from profits generated by foreign firms' subsidiaries in Ireland.

How did this affect the Irish balance of payments? Figure 72.2 illustrates this like no words possibly could. As FDI flows to Ireland dried up and Irish consumption fell due to falling property prices, rising unemployment and falling incomes, the previous trend in current account deficits was immediately broken. Seldom does one see such stark correlation backing up economic theory as the two curves showing falling GDP

and improvement in current account. FDI decreased, the Irish decreased import spending and thus the capital/financial account surplus started to decrease. This of course meant that the current account deficit decreased – and by 2011 became a surplus.
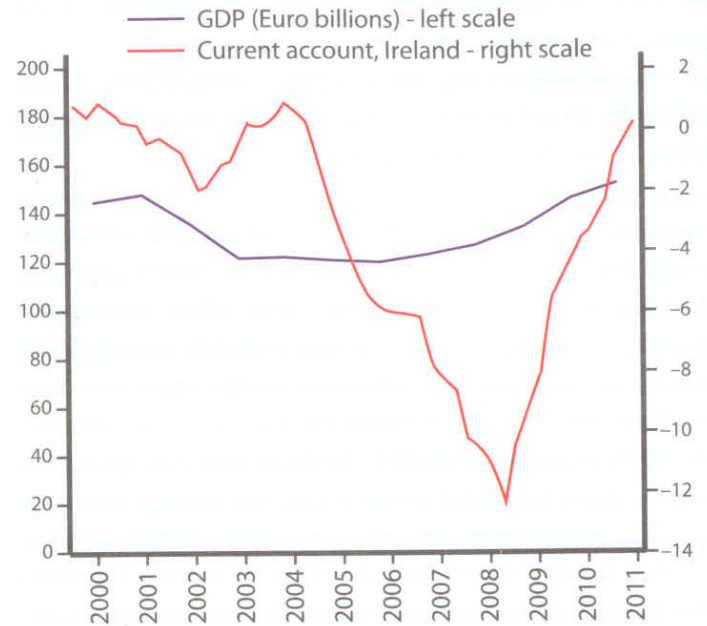
Figure 72.2 Irish growth and current account, 2004 – 2010

However ...

On the other side of the ~~fence~~...that's *Atlantic*, the US has run a current account deficit every year since 1991, shown in figure 72.3. Basically, foreign investors considered the risks smaller, and the returns greater, in investing in the vibrant and innovative US economy than at home. In other words, foreign capital flows have benefited the US economy by creating funds for investment which served to lower unemployment and increase GDP considerably during the 1990s. A current account deficit allows a country to enjoy greater consumption than production – even though it might be on borrowed money. If the deficit is relatively short-lived, a few years or so, then there would be little economic damage – quite the opposite if the inflows in capital account are partially used for investment.

Even *long run* current account deficits may have relatively little impact on the domestic economy. A good many countries have run current account deficits for many years in a row without any seeming ill effects. A main reason is that while a current account deficit of hundreds of millions of dollars/EUROS sounds ominous, it must be related to the countries' fundamental ability to pay the money back – i.e. the size of national income.

1   *Worldwide Capital Shares and Rates of Return to Corporate Capital: Evidence from U.S. Multinationals*, Mihir A. Desai, Harvard University and NBER, November 2001

2   Purchasing power parity – see Chapter 81.

3   *Taxation and foreign direct investment in Ireland*, Brendan Walsh, 2001. This is an update of Stabilisation and Adjustment in a Small, Open Economy: Ireland, 1979–95; Oxford Review of Economic Policy 12, 3. (October): 74–86.)

4   *National Income and Expenditure 2002*, Irish Central Statistics Office, http://www.cso.ie and www.bullfax.com *Ireland's Economic Crisis: A brief summary*, 12/08/2010.

511

In the case of the US, which has now broken every record in the book concerning current account deficits, the current account deficit bottomed out in 2006 at around USD875 billion – 7% of GDP. In testimony before the US Budget Committee in 2007, the Peterson Institute for International Economics put forward the view that the deficit was '…clearly unsustainable'[5]. It was pointed out that any form of 'overspending' means that one has to in some way finance the debt. To uphold such a current account deficit the US had to attract foreign capital to the tune of close to USD4 billion … every working day!

This was, and is, in all likelihood unsustainable. Trace your ruler across Figure 72.3 and think about the correlative issues you have read about in this chapter. The shaded 'pillars' represent US recessions which clearly indicate that when US income fell the current account improved – however briefly! Again, not much correlation between the exchange rate and current account over the *entire* period but after 2001/'02 the massive current account deficits clearly take a toll on the US dollar exchange rate. Huge increases in government spending under president Bush II 2002 resulted in some of the largest deficits in US history. Ultimately this would have to affect US citizens' demand for imports and foreign demand for exports – by 2006 the current account was improving and three years later the current account was back to levels not seen for almost 10 years. The moral of the story here is that there is indeed a relationship between the exchange rate and the balance of payments – but it is extremely difficult to accurately predict 'what causes what and when'.

5 *The Current Account Deficit and the US Economy*, by C. Fred Bergsten, Peterson Institute for International Economics, Testimony before the Budget Committee of the United States Senate February 1, 2007 (See http://www.iie.com/publications/papers/paper.cfm?ResearchID=705)
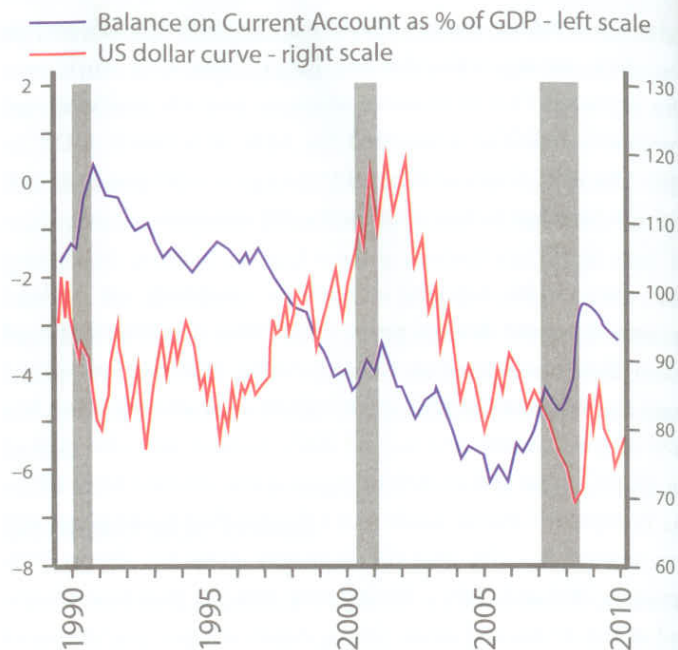
Figure 72.3 US current account and US dollar (trade weighted), 1990 to 2010

POP QUIZ 72.1

Balance of Payments

1. A firm in Argentina sells $US250,000 worth of beef to Canada and then the firm buys $US50,000 worth of machines for slaughterhouses from Brazil. What is the *net effect* on each country's balance of payments due to these two transactions?

2. What is the result of a country's capital/financial account balance minus the invisible trade balance?

3. We have the following figures in the balance of payments (billions) for a country; exports = €340, imports = €380, net invisibles = + €50, net flows in capital/financial account = - €120. What is the current account balance?

4. "The US current account deficit is increasing all the time and it's China's fault – they are stealing our jobs." This oft-heard statement provides you with a clue as to the state of China's current account. Explain.

5. How would a visit to Thailand by Italian IB students affect both countries' balance of payments?

6. Explain how a capital account surplus in Kenya might create jobs and increase national income.

7. Why might a country have a persistent current account deficit for many years?

Correcting a current account deficit

While there is no general *rule* stating that a current account deficit is harmful to the economy, many countries have considered this to be a form of balance of payments disequilibrium and thus something to be dealt with using various policies. When a country shows an alarming current account deficit – and there is little agreement as to what 'alarming' means – there are a number of economic policies which might be used to alleviate the situation.

Short run policies, invariably government/central bank intervention impact on the market in such a way as to lower the ratio of the price of domestic goods to imported goods. One possibility is to 'manage' the *exchange rate* in order to make domestic goods cheaper for foreigners. Another method is of course *protectionism*; if domestic goods are subsidised and/or tariffs levied on imports then import spending will fall and improve the current account. Contractionary policies aimed at reducing overall demand – e.g. aggregate demand – serve to lower overall expenditure, which includes imported goods.

Long run policies commonly focus on enhancing domestic competitive abilities, i.e. increasing productivity, increasing R&D and introducing innovation, improvements in quality and the like. These are looked at under *supply-side policies* further on.

Managing the exchange rate

A country operating under a managed currency regime, e.g. a pegged exchange rate system, can devalue its currency in order to alleviate a persistent current account deficit. By pegging the home currency at a lower exchange rate, the relative price of domestic goods falls. (Recall that a lower price does not necessarily mean more revenue! A fall in export prices might actually mean that while export *volume* increases, export *revenue* does not. HL, see **Marshall-Lerner condition** further on.)

The same basic outcome is possible in a floating exchange rate regime – which would of course then be a *depreciation* rather than a devaluation of the currency. The central bank of a country operating under a floating exchange rate system could sell home currency on the foreign exchange market in order to increase supply and lower the exchange rate.

Protectionism – expenditure *switching* policies

Recall that protectionism can be broadly defined as any policy where the ratio of the price of domestic goods to imported goods falls, i.e. imports become relatively more expensive. Devaluation and interventionist depreciation both serve to decrease the home country's demand for imports, which means that demand for goods is *diverted* from imports towards domestic goods. This *substitution effect* is known as **expenditure switching**. Another method – which is actually illegal for WTO members to use[6] – is the use of tariffs and quotas to limit imports and force home citizens to consume domestic goods.

> **Definition: 'Expenditure-switching policies'**
> Policies which divert – substitute – domestic expenditure away from imports towards domestically produced goods are **expenditure-switching policies.** Trade barriers and/or intentionally lowering the exchange rate (devaluation or depreciation) are examples of such policies.

However ...

While devaluation and tariffs work well in countries which have a high propensity to import, the former is a bit like opening eggs with a katana (Japanese long-sword) while the latter is more like using a sledgehammer. In other words, countries increasingly consider such methods a trifle heavy-handed, and while the egg won't fight back, expenditure switching policies frequently invite *retaliatory protectionism* and/or reciprocal devaluations.

TRADE

---

6 But of course, anti-dumping tariffs are allowed.

Reducing aggregate demand – expenditure reducing policies

Household spending is greatly influenced by income, which of course means that a degree of income will be spent on imports. An overall reduction in aggregate demand will lower incomes and *reduce imports*. Therefore, deflationary fiscal and monetary policies can be implemented in order to adjust a current account deficit by reducing imports – this is an **expenditure-reducing policy**. (HL will recall from Chapter 47 that the marginal propensity to import – MPM – is defined as the change in imports over a change in income; $\Delta M / \Delta Y = MPM$. The higher the MPM, the larger the effect on current account will be due to an expenditure reducing policy.)

Definition: 'Expenditure-reducing policies'

Contractionary policies such as increased interest rates and/or decreased government spending will cause a decrease in aggregate demand and a general **reduction in expenditure** (national income). Overall lower expenditure levels will also decrease the demand for imports.

It is worth noting that there is also an expenditure-switching element in deflating an economy. Lower relative inflation might cause home citizens to substitute imports with domestic goods. It is also possible that a reduction in aggregate demand lowers inflation (relative to trade partners) and thus further improves the current account by increasing demand for exports.

However ...

By lowering aggregate demand in order to improve the current account, there will be secondary negative effects on *employment and growth* in the economy. The use of deflationary policies to correct a current account deficit once again illustrates conflicts which arise in macro issues, i.e. the possibility of a **trade-off** in accomplishing both macro goals of high growth/employment and external balance.

Supply-side policies

There are a number of policies which will increase long run aggregate supply by increasing the ability and propensity of firms to produce and labourers to supply. (See Chapters 60 to

62.) By reducing labour costs, adding to labour skills (human capital), creating incentives for investment in technology and generally increasing productivity, a country can increase its international competitiveness. This would ultimately increase exports and also divert some spending towards domestic goods rather than imports.

However ...

The problem is that such supply-side policies commonly take several years to implement and even longer before the effects are visible in the balance of payments.

POP QUIZ 72.2

Balance of Payments and Exchange Rates

1. How will the current account be affected if a country supports its exchange rate so that it is above the free market equilibrium?

2. Why might a country raise interest rates in order to sustain a current account deficit?

3. What are the possible long run benefits of a current account surplus?

4. Which measures imposed by a country to reduce a current account deficit would be described as 'expenditure switching'?

5. Why might an increase in personal income tax in an economy improve the current account in balance of payments?

6. Say a government wished to improve a current account deficit immediately; would you recommend devaluation or supply-side policies? Explain your choice.

Effects of a long run current account surplus

A current account surplus means that there is a net outflow in capital/financial account, i.e. the home country's net foreign assets have increased. There will be a number of gains for such a net *creditor* nation:

- The foreign assets can be viewed as another form of saving for the home country which will enable increased **future consumption**.

- Capital will flow to countries with a higher rate of return than the home country. This enhances **resource allocation and increases profits** for domestic firms.

- The increase in foreign holdings will in time **generate income** in the forms of profits, interest received, and dividends. (Inflows in current account.)

In spite of these there are several **possible disadvantages** in having sustained current account surpluses:

- **Current consumption** possibilities for the home country decrease as resources are diverted abroad.

- A current account surplus means that there is a degree of **diverting investment** from the domestic to the foreign market. This could lead to a loss of jobs (yet this is highly contentious), skills and technology gains.

- There will be a degree of **tax loss**, as a portion of tax bases – investment, output and wages – will be taxed outside the home country.

- There is also a **political element** in running continuous current account surpluses for many years. This was a main ingredient in trade friction between the US and Japan during the 1980s – and between the US and China during the 2000s. Guess which of these countries were running large current account surpluses? (See footnote.[7])

General external equilibrium

The link between equilibrium in the balance of payments and equilibrium in the exchange rate is very strong. The two concepts are technically two sides of the same coin; when there is no pressure on the exchange rate to change, there will theoretically be a balance of payments equilibrium. If the value of exports – expressed in a currency – is equal to the value of imports, then there will be no excess demand or supply for the currency. Should the home economy become more efficient and thus more internationally competitive, exports will rise and so too will the demand for the home currency. Naturally, the opposite holds true.

However …

As we have seen, in reality a currency can easily appreciate for lengthy periods while consistent current account deficits rise. This is primarily due to investment and speculative demand for the home currency. We have looked at the former so now let us have a look at how general disequilibrium can arise due to speculation in the case study below.

CASE STUDY

SOUTH EAST ASIAN CRISIS OF 1997/'98

By 1997, the situation in Asia was basically the following: The Tiger economies (Thailand, Taiwan, South Korea, Hong Kong, Malaysia, Indonesia, Singapore) were growing at an annualised rate of between 5 – 12%. Massive FDI was pouring in from OECD countries, primarily EU countries and Japan, (USA at a lower level) as the successful resolution of the Mexican economic crisis in 1995 plus the fall/death of communism (and thus the threat of a 'take over' of private property) made investment in LDCs safe. The 3rd world, or developing nations, was re-dubbed 'Emerging Markets' by the Marketing Babble Section of Very Large American Investment Firms. The flow of private funds from Europe and Japan to developing nations increased by about 400% during the period 1990 to 1997!

---

7   China and Japan. The all too common view was that China has been growing at the United States' expense. See a good high level debate at http://www.cfr.org/publication/11631/.

The Bubble-Baht:

This is how funds got from Japan/Europe to Thailand (where the crisis started):

A) A Japanese bank would make a loan to a Thai finance company – which is an institution that acts as a 'go-between' for foreign capital.

B) The finance company then had Yen which it would lend to a local real estate developer (= construction firm). However, the developer needed baht, (= Thai currency) not yen, to pay the construction crew/land…etc. The developer would go to the foreign exchange and buy baht.

C) Result; the demand for the baht increased and thus the baht appreciated.

D) Now, the central bank of Thailand had committed itself to maintaining a stable (= pegged) exchange rate to the USD. To offset the increase in demand for the baht, the central bank had to increase the supply of baht – i.e. print more baht. The result now was that the Thai central bank had larger currency reserves and the money supply increased.

E) Credit was now easy, investment soared and foreign investment and monies (speculative now!) continued to pour in.

Why didn't the Thai government simply abandon the peg to the USD and let the baht assume a higher exchange rate? Answer: because this would have damaged Thai exports.

F) Thailand started to see a larger and increasing current account deficit –remember, FDI and other inflows of currency on capital account necessarily entail a deficit on current account. Is a large current account deficit necessarily evil? It depends on what the inflow on capital account is used for ….

G) In the case of Thailand (and indeed Indonesia) a good deal of the foreign funds were leant to high-risk projects with very questionable securities. Basically, the finance company that borrowed the yen in the first place, was nothing more than an institution run by a relative/friend of the government! This is where 'cronycapitalism' comes in.

H) Foreign investors and lenders (= banks) to the finance company, would not ask for the normal securities of such loans – the company was, after all, run by the nephew/friend/schoolmate of the minister of finance. How risky could the loan be?!

I) So, the nephew would get a low interest loan from abroad and then lend the money to his friend the real estate developer who's planning a new Mega Tower. If the Tower was successful, everybody wins. If the Tower is a total disaster (= no renters/ buyers) then the taxpayer will lose as the finance company will be bailed out by Uncle Finance Minister: 'Heads, I win; tails, you taxpayers lose'. Now we have a speculative bubble economy!

J) Ultimately, firms started incurring losses; finance companies and real estate companies started to go bankrupt. Foreign investor confidence started to decline and less money flowed in. The baht started to lose value.

K) The government had two alternatives:

1) Protect the baht by reducing the supply of baht and raising interest rates. The Thai government waited (as this would cost them foreign reserves and the interest rise would hurt the already ailing economy) and the market started to regard the baht as soon to leave the peg to the USD. But as long as the exchange rate remained, one could borrow baht and in ANTICIPATION of the baht being devalued! That's what the market did; borrowed baht and exchanged these immediately for USDs, thereby increasing the supply of baht even more!

2) Let the currency float and accept a lower exchange rate ... and the Thai government still waited – as this would have meant that hundreds of domestic firms already in trouble would have to pay far more baht to service foreign debt in US dollars! So the government defended the baht – and failed. Ultimately, the peg to the US dollar was abandoned and the baht fell by 50% in a few months taking the Indonesian rupiah, Malaysian ringgit, Korean won and Hong Kong dollar with it.

The devaluation thus caused a string of devaluations in South East Asia. As the Thai foreign currency reserves – the 'war chest' – were depleted, the only option left to keep an element of foreign confidence in the baht was to raise the interest rates and reduce the supply of baht. This further aggravated an economy under severe recessionary pressures.

This was the 1997 'meltdown' of South East Asian economies. It was basically caused by speculation and resulting bad debt, and furthered by lack of liquidity, high interest rates and economies in deep recession. Again, we are dealing with economic basics; expectations and confidence. The financial crisis can be seen as a negative self-reinforcing loop – a vicious circle; loss of confidence in Asian economies → currency values plummet, interest rates rise and Asian economies suffer from recessionary pressure → financial problems for Asian banks/firms/ households → loss of confidence in the Asian economy ... etc.

Answers

1. Since there is a current account deficit, it must be made up for by a capital/financial account surplus. The large net inflow in FDI explains this. There is also the possibility of speculative inflows in 'Net other investment'.

2. No, it's just under 1.5% of GDP which is historically low.

3. Highly unlikely since there is a net income outflow (debit) in current account.

4. If exports continue to exceed imports, it is likely that the supply of the Home currency is increasing. This can cause a depreciation of the exchange rate.

5. A current account deficit would put downward pressure on the exchange rate. In order for the central bank to retain the peg towards the foreign currency it might raise interest rates and/or intervene on the Forex market by support-purchasing its own currency.

6. I would say so since almost 15% of GDP comes from export revenue. For comparison I point out the corresponding figure for Japan has been between 13 and 16% for over 20 years.

7. Decreased since there is a net inflow in reserve assets which means a *decrease* in net foreign assets, e.g. in the foreign reserves held by the central bank.

Summary & revision

1. A persistent **current account deficit** can result in:

   a. A *depreciation* of the exchange rate (large imports increase the supply of the Home currency)

   b. Large *foreign debt* (due to inflows in the financial account)

   c. Higher interest rates (as increased foreign debt means the debtor nation must offer a 'risk premium' to foreign creditors)

   d. Higher interest rates can have a contractionary effect on the Home economy

   e. Speculative bubbles can arise due to the inflows on financial account

2. A current account deficit might be **corrected by:**

   a. *Devaluing/depreciating* the home currency (making exports cheaper and imports dearer)

   b. *Expenditure switching* policies, e.g. protectionism (tariffs, quotas and subsidies lower imports)

   c. *Expenditure reducing* policies such as raising interest rates to curb AD and thus import spending

   d. *Supply-side policies* (increased international competitiveness – this is not a short run solution since supply-side policies take considerable time to kick in)

3. **Positive effects** of a long run **current account surplus** include:

   a. *Higher consumption* for the Home economy

   b. *Improved resource allocation* for domestic firms as current account inflows enable foreign investment

   c. Increased *future profits* and income for the Home economy as investment abroad creates repatriated profits

4. **Negative effects** of a long run **current account surplus:**

   a. Diverting resources abroad means a loss of *current consumption* at Home

   b. *Investment is diverted* from the Home economy

   c. *Possible tax losses* due to lower investment levels at home

   d. *Political issues* arising as trade partners battle with current account deficits

73. Marshall-Lerner Condition and J-curve

Key concepts: HL extensions

- Marshall-Lerner condition
- J-curve effect

This HL section looks into the effects of devaluation on exports and imports. I beg you once again to keep in mind the distinctions between export **price**; export **volume**; and export **revenue**.

Marshall-Lerner condition

The effects of a devaluation appear relatively straightforward; a devaluation of the currency will lower the price of exports and increase the price of imports, which in turn will increase the volume of exports and decrease the volume of imports. However, the degree to which export *revenue* and import *spending* is affected – and thus the current account – depends on the price elasticity of demand for exports and imports.

Choose a pair of baskets...

The series of diagrams (A to D) in figure 73.1 show a number of possibilities arising from devaluation. Most important note here: the diagrams are simply a method of explaining how different elasticities will affect export revenue and import spending. I have taken great liberty – known as cheating – in my use of economic concepts by stating that a change in the exchange rate will cause a movement along the demand curve for both exports and imports. The connection between the exchange rate (which is in effect Home's price of exports and imports relative to trade partners) is in fact a good deal more complex than this. I am using the demand curves below ONLY TO EXPLAIN/ILLUSTRATE the effects of price elasticity of demand on exports and imports. I recommend that you do NOT use them in internal assessment or exams.
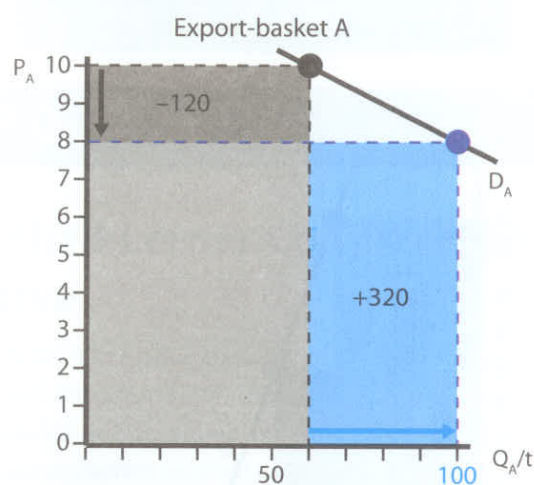
Net increase in export revenue = +200
(Effect on current account: +200.)
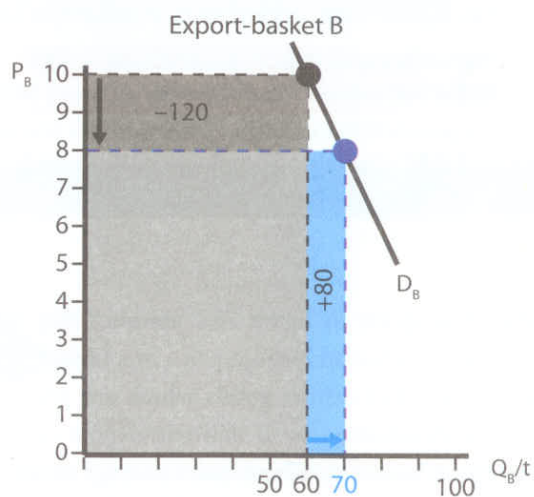
Net decrease in export revenue = -40
(Effect on current account: -40.)

Figure 73.1 A&B Devaluation/depreciation and possible effects on current account

HIGER LEVEL

TRADE

Export-basket C
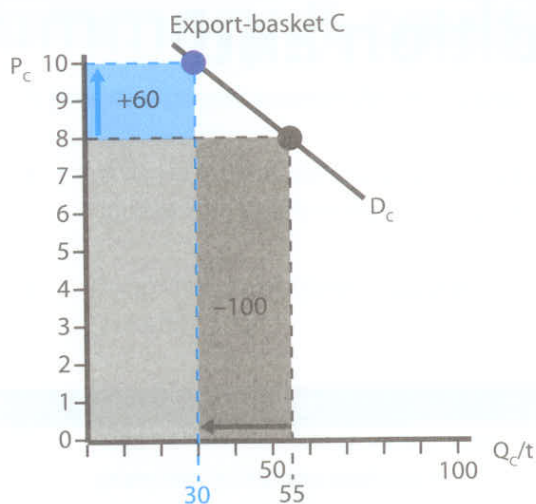
Net decrease in import expenditure = –40
(Effect on current account: +40.)

Export-basket D
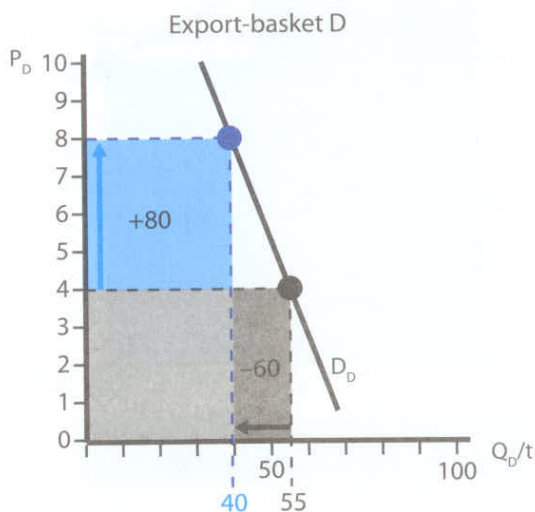
Net increase in import expenditure = +20.
(Effect on current account: –20.)

Figure 73.1 C&D Devaluation/depreciation and possible effects on current account

Note that each diagram in figure 73.1 assumes that 'price' is not for a single good but an *average* price of a basket of export and import goods. For export goods which are price elastic, the result of a devaluation is to increase export revenue, as shown for **export-basket A**, where lower average export prices result in a loss of export revenue of $120 and a gain of $320. The relatively high price elasticity of demand for this country's export goods results in a *net increase in export revenue of 200*. In raising the price of imports, devaluation will also affect total import spending. Assuming that imports are also relatively price elastic, as in **import-basket C**, there will be a net decrease in import spending of $40.

Summing up, if the country in question has an average price elasticity of demand for exports as in basket A, and average price elasticity of demand for imports as in basket C, there will be a net increase in the current account improvement of 240. However, if the combinations of average price elasticities for export- and import-baskets change then there are other possibilities:

- Export basket A and import basket D; current account balance *increases* by 180

- Export basket B and import basket D; current account balance *decreases* by 60

- Export basket B and import basket C; *no change* to current account balance

These examples show that for a devaluation to be considered successful, certain conditions of price elasticities must be upheld. It is possible for the current account to improve even when either exports or imports are demand inelastic. The **Marshall-Lerner condition** summarises this in stating that *devaluation will improve the current account as long as the sum of the price elasticity of demand for exports plus the price elasticity of demand for imports is greater than one*. Conversely, if the sum of price elasticities of demand for exports and imports is less than one, a *revaluation* of the currency will improve the current account.

Definition: 'Marshall-Lerner condition'

If the sum of the price elasticities of demand for exports and imports is greater than one, a devaluation will improve current account in the balance of payments. Thus:

Devaluation → $PED_X + PED_M > 1$ → current account improves

Revaluation → $PED_X + PED_M < 1$ → current account improves

J-curve effect

The conclusion that a devaluation leads to an increase in net export revenue is generally valid. However, depending on how quickly households, importers and exporters respond to the *de facto* change in relative prices, *there might initially be a worsening* in the current account. There are three main reasons why this might be the case:

1. The Marshall-Lerner condition may not be upheld as the price elasticities for exports and imports are likely to be **relatively inelastic in the short run**. It will take time for firms importing raw materials and components to find alternative sources, so a devaluation will in fact worsen the balance of payments as firms in the short run continue to purchase imported factors of production at higher import prices.

2. Firms are **locked into contracts** for months in advance. An importer with a contract in foreign currency to be paid 6 months down the road will have to pay more to fulfil the contract after a devaluation. Similarly, exporters contracted in foreign currency will receive less in home currency. Both cases will cause the current account to worsen.

3. It takes time for both firms and consumers to adjust to the fact that relative prices have changed and thereby **change expenditure habits**.

Figure 73.2 J-curve effect on current account

Figure 73.2 illustrates the J-curve effect of a devaluation:

- t_0 to t_1: A current account deficit is countered by a *devaluation at* t_0 whereupon there is an initial worsening up to point t_1 as households and firms have difficulties in adjusting to the new exchange rate in the short run.

- t_1 to t_2: Ultimately, domestic firms will find alternative (lower cost) sources of factor inputs and long run commitments (contracts) will end, households will alter spending habits more in favour of domestic

goods, and foreign trade partners will increase their purchases of goods and services of the country which has devaluated. Taken together, import expenditure is falling and export revenue is rising so the current account starts to improve. At t_2 the current account is back to pre-devaluation level.

- t_2 **and beyond**: Beyond t_2 the balance of payments shows a smaller current account deficit than pre-devaluation and at t_3 the balance of payments is in perfect equilibrium. Beyond t_3 the current account moves into surplus. (The portion of the net export curve between points t_1 and t_3 show the 'J' portion of the J-curve.)

Marshall-Lerner Condition and J-curve

1. A country's price elasticity of demand for imports is 1.3 and for exports 0.5. The country operates under a fixed exchange rate regime and desperately wants to correct a severe current account deficit. What do you recommend?

2. Is upholding the Marshall-Lerner condition an example of expenditure-switching or expenditure-reducing?

3. Is it possible to have a J-curve effect even in the short run when the Marshall-Lerner condition is upheld?

There is some evidence that depreciation might in fact not have a great effect on current account. Have a quick look back to Chapter 69 and Figure 69.3. The Chinese Yuan appreciated by some 20% between 2005 and 2008. The effect on the US current account? The US trade deficit with China (note; solely with China and not with other trade nations) actually *grew* by the same percentage.[1]

1 *On the Yuan, be careful what you wish for*, Business Week, 23 September 2010

Summary & revision

HIGER LEVEL

1. The **Marshall-Lerner condition** states that a devaluation/ depreciation of the exchange rate will have a positive effect on current account in balance of payments if the sum of PED for exports and the PED for imports is greater than $|1|$.

2. The **J-curve effect** builds on the concept of price elasticities for exports and imports, stating that a devaluation might lead to a worsening of current account in the short run. This is due to inelastic short run demand for imports – it takes time for households to adjust their spending patterns and firms are locked into contracts.

TRADE

3.4

74. Forms of Economic Integration

Key concepts:

- Bilateral and multilateral trade agreements
- Trade blocs (REAs)
  - Free trade area
  - Customs union
  - Common market
- Monetary union (EMU)
  - Advantages of monetary union (EMU)
  - Disadvantages of monetary union (EMU)

Bilateral and multilateral trade agreements

When two countries agree on obliterating or lowering tariff and non-tariff barriers across a range of goods one speaks of a **bilateral** trade agreement – also known bilateral free trade agreements. There are – literally – hundreds of bilateral trade agreements in the world and they differ substantially. The main theme is that the signatory nations give each other *preferential* treatment in trade (lower/no tariffs and other barriers) that non-signatory nations will not have. In 2011 alone eight bilateral agreements were signed, ranging across the globe; Canada – Columbia, China – Costa Rica and India - Japan to name but three.

When three or more nations sign an agreement binding members to rules set down by a common contractual agreement, there is a **multilateral** trade agreement, the largest and most important of which is the WTO. Here too there is great variation as to

which trade barriers are affected and by how much, so I shall limit the discussion to an explanation of the WTO's (Chapter 63) most favoured nation clause (MFN) and overview of the 'stages' of economic integration.

> **Definition: 'Bilateral and multilateral trade agreements'**
> A **bilateral** trade agreement is between *two* countries and is designed to increase trade by way of lowering tariff and non-tariff barriers on a range of goods. A **multilateral** trade agreement is between *three or more* countries.

Most favoured nation (MFN) clause

The portal paragraph of the WTO states that member countries cannot discriminate between trade partners. Any special favour such as lowering tariffs for one country means it must be done for all other member countries. The WTO rules allow certain exceptions; preferential trade agreements and free trade agreements (FTAs, covered further on) are the main ones.

Overview of economic integration

Integration is more of a process than a final stage. Figure 74.1 illustrates six stages of increasing economic integration. I include a few examples of each stage, going into further depth in stages II – V further on.

TRADE

Figure 74.1 Stages of integration

Stage	I: Preferential trade agreement	II: Free trade agreement (FTA)	III: Customs union
Type of agreement	Reduced tariff levels for selected countries.	Zero tariffs on selected goods. Each country keeps its own tariffs towards non-members.	Zero tariffs on goods between members and a common tariff towards other countries.
Example	Lomé /Cotonou Agreement between the EU and African, Caribbean and Pacific (ACP) countries.	North American Free Trade Agreement – NAFTA	EEC (EU) prior to 1993.
Difference from previous stage	Tariffs are selectively lowered	Tariffs are removed rather than lowered	Setting of a common external tariff (CET)
Stage	IV: Common market	V: Economic & monetary union	VI: Full integration
Type of agreement	Free movement of goods, services, capital and labour between members.	Fixed exchange rates or single currency.	Harmonisation of tax laws, social systems, and federative/'supranational' state.
Example	European Union (EU)	European Monetary Union (EMU) in the EU	For example, the German and US federal system.
Difference from previous stage	The 'four freedoms' of movement above result in convergence in taxes, laws and social welfare systems	Convergence in monetary and fiscal policies. (16 EMU members of the EU 27)	Centralised economic, political and legislative power of government.

Trade blocs (regional trade agreements - RTA)

A most visible aspect of the on-going process of economic integration in the world today is the many different trading blocs which have arisen over the past 30 to 40 years. All of the 153 WTO members have at least one REA – the only exception being Mongolia – and the average number of agreements each country is a signatory to is 13.[1] The WTO had 319 **Regional Trade Agreements** (RTAs) registered in 2011 and intra-RTA trade (= trade within an RTA) comprises just over 50% of global trade. The EU and NAFTA alone account for 62% of all exports in the world.[2]

Free trade area

When a group of countries sign an agreement whereby tariffs and quotas are removed between signatories, i.e. member countries, while keeping individual tariffs/quotas towards non-members in place, **a free trade area** has been constructed. While seemingly a straightforward method to encourage and increase trade, the proliferation of FTAs has resulted in a number of criticisms of the subsequent *exclusion of non-members* – notably from developing countries which are frequently disadvantaged.

Definition: 'Free trade area' (FTA)

A **free trade area** arises when a group of countries remove tariffs and quotas between themselves, while retaining the right to set tariffs/quotas towards non-members.

---

1   World Trade Report 2011, pages 60 – 64 and WTO at http://www.wto.org/english/tratop_e/region_e/region_e.htm

2   World Trade Report 2011, page 68

TRADE

Figure 74.2 I and II show the progression of three countries, A, B and C. (To simplify, we assume that the same tariff level is applied to *all* goods entering each country.) Notice that while all three countries have different tariff levels towards each other and the 'outside world', each country has the same rate for all countries. Country A has a tariff of 7% on goods coming from B, C and other countries; Country B has a 9% tariff and C a 5% tariff. This is in keeping with one of the key rules of the *World Trade Organisation's* (WTO) policy of non-discrimination, notably manifested in the MFN – the most favoured nation clause. Thus, in setting a 5% tariff on goods from Country A, Country C has had to set a maximum 5% tariff towards all other countries.

I: Trade in accordance with MFN rules
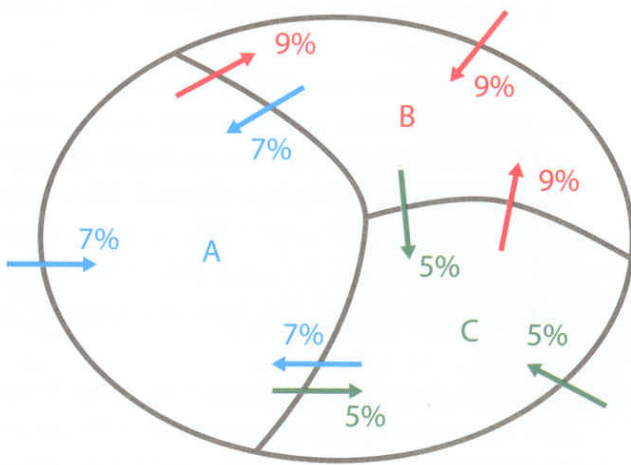
II: Free trade area – FTA
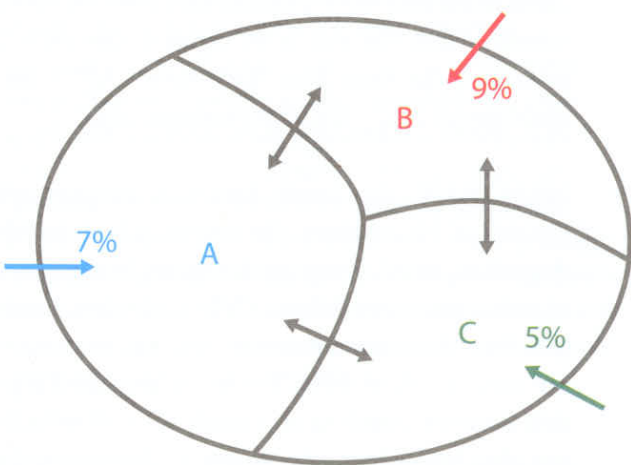
Figure 74.2 Constructing a FTA

In forming a free trade area, the three countries remove tariffs and quotas between themselves – but retain the right to set individual external tariffs towards non-members. (Note that the WTO rules allow for such exceptions from the general rule of

non-discrimination.[3]) Figure 74.2 II illustrates how trade flows freely between members yet each of the three FTA members keep the same external tariff.

Implications of a FTA

There are several implications of removing trade barriers from members:

- Firstly, there will be an incentive to trade more within the FTA and as a result **increase intra-FTA trade**.

- Secondly, governments of member countries are likely to experience a **reduction in tariff revenues** when overall trade flows increasingly are between non-tariff countries within the FTA.

- Thirdly, there will be an **increase in economic welfare** amongst members as prices of goods fall and efficiency losses are reduced.

- Fourthly, a larger tariff-free market can create long run **dynamic benefits** arising from larger markets and increased competition, such as benefits of scale and spread of technology.

- Fifthly, countries outside the FTA might in fact be more efficient in the production of certain goods but these **non-members will be disadvantaged by the tariffs**, which would results in a degree of misallocation. (See "Trade diversion" further on.)

Local content rules in a FTA

Ever used a friend's membership card in a video rental chain to get a better deal on renting a film? This is similar to the problem which faces members of a FTA; there is an incentive for producers outside the FTA (see figure 74.2 II) to avoid the highest tariffs by bringing in goods via **Country C** (with the lowest 'external tariffs') and then re-exporting to **Countries A and B** (also known as *trade deflection*).

3 This is in all probability due to the difficulties arising in a rather tricky issue: If Countries A and B create some form of political/legal entity, e.g. a 'United States of A-B', then trade becomes a domestic issue and not subject to WTO regulation. The tricky issue is how to define a political or legal entity?! The WTO leaves this alone by allowing countries to establish FTAs which are, in fact, highly discriminatory towards other WTO countries.

Say that cars produced outside the FTA are destined for **Country B** but are brought in via **Country C** in order to pay 5% tariffs rather than 9%. A $US15,000 car would then cost $US15,750 in **Country B** rather than $US16,350 when paying the full 9% tariff in Country B (discounting additional profit margins and administrative/transport costs). **Country C** would gain tariff revenue at **Country B**'s expense – Country B has in reality 'exported a tax base'! The response by a FTA member, in this case Country B, is to put a *tariff on goods which are re-exported* – commonly so that the final price of the good equals what would have been the case if the normal tariff rate of the destination country had been applied. Thus, the solution in this example is to put a re-export tariff of 3.8% on all cars coming into Country C from Country B. ($US15,750 $\times$ 1.038 = $US16,350.)

However...

Recall that goods are tariff-free if they are produced *within* the FTA! Non-FTA firms will respond to re-export taxes by setting up simple assembly plants in Country C and then importing the cars in pieces and large components. The car is then 'manufactured' in Country C and avoids re-export duties when exported to Countries A and B. Thus the FTA will have to implement yet another set of rules, **local content rules**, which set a minimum requirement on the value-added of final goods within the FTA. For example, NAFTA (see below) has a requirement that 62.5% of the content in terms of value-added of any automobile sold in the US from Mexico must come from within the FTA.[4]

Also, as free trade areas become increasingly common, there is talk of 'fortress Europe', 'walled-in North America' etc. When countries join RTAs there is a risk of international trade in fact becoming regional, with increasingly complex bilateral trade agreements between RTAs becoming the norm. There is increasing debate as to whether RTAs act as 'hubs' in creating global free trade or if they serve to undermine the intentions of the WTO in creating truly free trade in the world. Part of the discussion inevitably deals with foreign direct investment, as exporting countries which face regional trade barriers have increasingly responded by exporting factories – i.e. *foreign direct investment* – rather than the goods themselves in order to tunnel under REA tariff walls.

A Little Depth

Other Regional Economic Associations (REAs)

- **APEC** – Asia-Pacific Economic Cooperation: One of the largest RTAs in the world and potentially the largest FTA, comprising 21 member countries – from wealthy countries such as Japan, US, Canada and Australia to developing countries such as Chile, Peru, Indonesia and Vietnam. As yet APEC has not totally dismantled tariffs and quotas, but the aim is to have made the step to a fully-fledged FTA by 2015. The countries in APEC will together account for some 54% of world output and 44% of world trade. (See http://www.apecsec.org)

- **ASEAN** – Association of South East Asian Nations: This 10 member organisation overlaps APEC as many of the members belong to both. ASEAN is primarily comprised of developing countries – Cambodia, Indonesia, Thailand and Vietnam to name a few. Tariff levels have been preferentially lowered between members and the aim is to create a complete free trade area, the Asian Free Trade Area – AFTA, for all members by 2015. (See http://aseansec.org)

- **MERCOSUR**: This South American customs union consists of 10 countries; the founding four members Argentina, Brazil, Uruguay, Paraguay; Venezuela; and associate members Bolivia, Chile, Columbia, Ecuador and Peru. A common external tariff was implemented for most goods in 1995. Trade has increased a great deal between members as a result of tariff reductions but also highlighted the results of decreasing trade between members and non-members. (See http://www.mercosur.org)

---

4 NAFTA articles, article 403 at http://www.nafta-sec-alena.org/ DefaultSite/legal/index_e.aspx?articleid=116#A403

TRADE

CASE STUDY

THE NORTH AMERICAN FREE TRADE AGREEMENT

There are few examples which highlight the issues of REAs in starker colours than the North American Free Trade Agreement1, between Mexico, USA and Canada which was created in 1994. The area has over 360 million citizens and over $US1 trillion in GDP – both figures roughly the same as the EU's. As for the level of economic integration, tariffs between Canada and the US were successively removed by 2000, but complete free trade with Mexico did not take place until 2009 – primarily due to the two northern members 'fears of their agricultural markets being flooded by cheap Mexican produce'.

The creation and results of NAFTA are under continuous debate. Initial – and continuing – opposition to the treaty came from both left and right, environmentalists and 'smoke stack' industries, and even pro- and anti-globalisation groups. Perhaps the most memorable criticism was put forward in 1993 by then presidential hopeful Ross Perot, who claimed that the free trade agreement would result in a "…giant sucking sound" as jobs left the US in favour of far cheaper labour in Mexico. In addition to the loss of jobs, the poor of Mexico would inevitably be exploited while poor environmental standards there would attract high-polluting firms in an environmental race to the bottom. Proponents stated pretty much the opposite: jobs would instead be created in the US and Canada while consumers would benefit greatly as each member country exploited comparative advantages. Workers in Mexico would benefit from increased incomes and improved working conditions and overall intra-FTA trade would increase economic growth and development while decreasing poverty levels. Inward flows of FDI to Mexico would enable better use of resources and create positive technological and knowledge spin-off effects.

The outcome of the first decade is a mixed bag indeed and figures can support both sides pretty much equally. Proponents of NAFTA point to Mexico's 225% increase in exports and that this accounted for more than 50% of Mexico's real GDP growth during the first 10 years.

Trade between the countries has more than doubled in the same period, from $US306 billion to $US621 billion. The 'giant sucking sound' was never heard in America, as US exports to Mexico and Canada went from $US142 billion to $US 263 billion and US manufacturing output increased by 44% from 1994 to 2000.

In addition, Americans in manufacturing increased their real wages at twice the rate during NAFTA's first 10 years than they had during the preceding decade. Mexico 's real GDP per capita would have been an estimated 4-5% lower without NAFTA by 2002 and there is little evidence of any adverse effects on Mexican working conditions or wages – in fact, wages are considerably higher in sectors exposed to export markets.

On the other hand, point out NAFTA 's detractors, there has been no net addition of US jobs to speak of and in fact there has been a net loss of 110,000 US manufacturing jobs per year between 1994 and 2000. And while growth in the US and Canada has been impressive, Mexico has had several periods of zero and even declining growth, notably in 1994/95 and again in 2001 – real wages in Mexico took over 8 years to recover from the mid-'90s economic crisis. American opponents of NAFTA point to the over 170% increase in FDI in Mexico and over 430% in Canada during the 1990s to show how firms are relocating production outside of the US which will cost more jobs in the long run. As for economic growth, both the US and Mexico have experienced widening gaps in income between the poorest and richest 10% of the population, with an estimated 15 to 25% of the increased inequality resulting directly from increased trade. Finally, there has been growing discontent amongst the rural poor in Mexico, as small scale farmers are marginalised due to the increase in agri-business orientated agriculture and large imports of American corn.

The plan of bringing about a Free Trade Area of the Americas (FTAA) encompassing NAFTA and 31 countries from South America and the Caribbean has been debated since 1994 and is still being debated. At the time of writing, April 2012, the official website no longer exists.

TRADE

Customs union

When a stage of economic integration is adopted by which a *common external tariff* (**CET**) is set for member countries, a **customs union** has been created. These unified tariff levels solve the problem of re-exports, since any one country in the union will have the same tariff on goods coming from non-member countries. The CET also creates a need for far greater convergence, since different tax rates on goods need to be harmonised across the board. Thus, one could say that while free trade areas are politically uncomplicated and administratively complex (due to re-export and rules of origin agreements), customs unions are administratively simpler but more politically complex.

The tariff effects of a customs union are shown in Figure 74.3 I and II. Positing that our fictitious FTA moves to this next step of economic integration, the three countries A, B and C agree on a *common outer tariff* towards non-members – here an average of the three tariff rates. This stage is best exemplified by the **EU** - the **EEC** prior to 1995 - which came into being in 1958; the 1957 **Treaty of Rome** spelt out the phased elimination of all tariffs and quotas within the union and identical such barriers outwards. This was completed by 1987. (See Case Study on the EU below.)

Figure 74.3 illustrates once again how *trade might both increase and be diverted*; Country C will see increased trade with fellow members – but possibly at the expense of diminished imports of lower cost goods from trade partners who previously met tariffs of 5% but now meet a 7% tariff. (HL will study this **trade creation** and **trade diversion** in greater depth.) A customs union is somewhat like an 'infant common market' or 'common market lite', in that all goods and services may be moved freely between member states free of tariffs and quotas.

Note that there is no textbook definition of a customs union or a common market. Yet a clear distinction can be made here; a customs union may be said to deal with the obliteration of **visible trade barriers** on goods and services but not on technical, environmental and other 'invisible' barriers. Another notable difference would be that a customs union would have **limited freedom of movement for capital and labour** while a common market would have total freedom of factors of production.

Common market

When countries agree on moving towards market integration encompassing not only visible trade barriers on goods, but all hindrances in the flow of goods, services, investment, labour and other factors, a **common market** is established. A common market is complex and entails far greater depth in integration; economic integration is added to by social, legal and political integration.

I: FTA – Individual external tariffs

II Customs union – Individual external tariffs – common external tariff
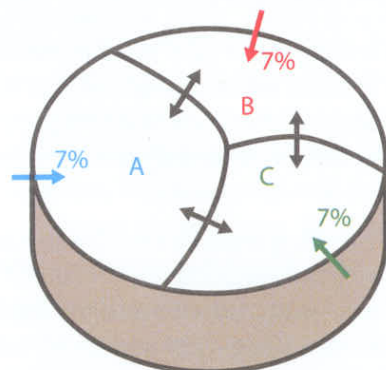
Figure 74.3 From a FTA to a customs union

> **Definition: 'Common market'**
> When countries agree on moving towards market integration encompassing not only visible trade barriers on goods, but all hindrances in the flow of goods, services, investment, labour and other factors, a **common market** is established. A common market is complex and entails far greater depth in integration; economic integration is added on to by social, legal and political integration. The purest example of a common market is the European Union.

TRADE

A common market = further integration

While there is – once again – no absolute definition etched in stone, a common market will ultimately have to implement several features of cooperative economic policies:

- **Factors of production** will be free to move across country borders, as will wages, profits and other incomes. A common market will therefore move towards broad scale convergence in **labour laws**, **competition legislation** and **financial rules**. For example, EU competition laws regulating mergers and monopolies are inserted into each country's national legislation and in fact take precedence over domestic legislation.

- **Non-tariff barriers** are removed, which necessitates far-reaching cooperation in technical/health/safety/ environmental standards. Product specifications, certification and technical standards are increasingly brought into a common line. The EU **mutual recognition rule** thus states that any product and/or production which is accepted in one EU country must be accepted in all.

- **Taxes** on labour, income, goods and profits must ultimately be coordinated in order to hinder tax evasion, avoid large scale arbitrage (= re-selling) of goods due to VAT differentials, and to limit regional unemployment arising when firms relocate to countries with more beneficial taxes. The EU has harmonised a good number of taxes. For example, VAT is paid only once, in the country of final consumption. From 2005, citizens working in another EU country will get their income figures registered by the home country's tax authorities.

The world's closest version to a common market, the EEC, was expanded during the 1970s, '80s and '90s, reaching 15 members by 1995 when the shift towards a common market was signified by a change of name to the European Union (EU). By this time a de facto common market (or single market) – in accordance with the four points above – had largely been in effect for two years. Figure 74.4 (I and II) illustrates how, to all intents and purposes, member countries now comprise a single unified market for all forms of economic activity.
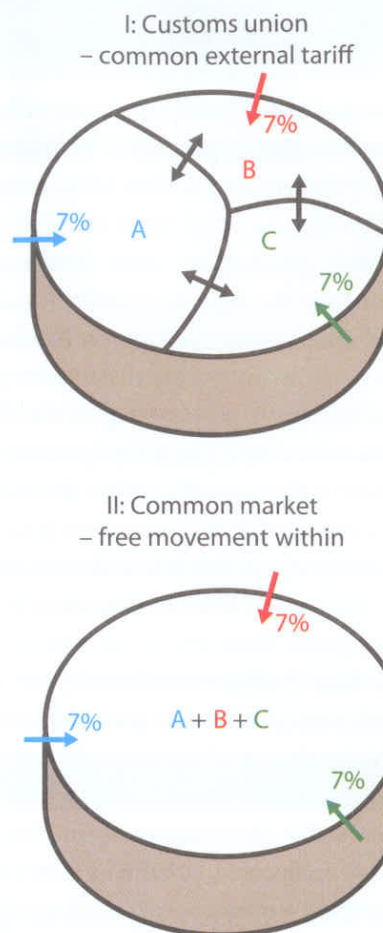
Figure 74.4 Constructing a common market

The impact on trade and economic integration is quite considerable. The EU has stated this most clearly in the "**Four Basic Freedoms**" of the free movement of goods, services, people and capital within the single market. **People** can move to other member countries to work, study or simply live. **Goods** can be bought and sold without hindrances across national borders as can **services**. **Firms** can operate freely across borders and **capital** can be transferred without restrictions.

The problem of sovereignty in further integration

"Hell hath no fury like a bureaucrat scorned".
Milton Friedman

Every person I know from an EU country has an indignant story to tell of how his or her country had to give up, adjust or somehow change a traditional good; beer in the Netherlands, sausage in Austria…or the technical definition of a cucumber in Portugal. Ever more interdependence between regional

TRADE

529

economic area members makes it increasingly necessary to have common policies and broad harmonisation of legislation in both economic and social areas. Quite naturally, this process has met with considerable suspicion and resistance from both citizens and policy makers alike within REA countries.[5]

The famous British philosopher John Locke claimed that legislators do not have the right to transfer legislative powers into the hands of others, since their power has been delegated (= entrusted) to them by the people. Resistance to integration arises when countries have to give up a degree of self-government and abide by new rules created by a foreign institution. At the centre of the debate is the issue of whether de-democratisation takes place; can a *supranational* (= members grant power to an outside institution) body in fact claim *democratic legitimacy*? When countries increase economic integration by establishing a RTA, increasing goods flows will create a need for rules on a variety of issues; from re-export tariffs and rules of origin in a FTA, coordinated taxes and labour laws in a customs union/ common market, to decreased freedom in setting monetary policies in a single currency area with a Central Bank which governs member states. Sovereignty is in this manner '... never lost, but only reallocated. ... When a state's sovereignty is reduced, the important question raised is where the sovereignty goes.'[6] Three areas illuminate a few core issues of countries' reluctance to cede (= give up) political sovereignty:

Legislation: Different laws evolve over time in different countries, and there will be resistance to giving up a degree of legislative power to 'supranational' institutions. The current best example of this is within the EU; when EU members agree to be governed by legislation put forward by the **EU Commission** (which is not elected by EU citizens) and undersigned by the EU Parliament (elected by about one third of EU citizens eligible to vote), member states have in fact given up a degree of self-rule. The degree of supranational legislation in the EU is considerable; EU law to a large extent has precedence over national laws in areas such as competition and labour – and any new legislation in member states must not contradict existing EU legislation. Perhaps nowhere has resistance to further

integration been stronger than in the case of the **European Monetary Union** (see further on) and the on-going attempts to draft an *EU constitution* regulating the division of sovereignty between member states and the EU Parliament/Commission.

Taxation: When countries form a REA there is a tendency for cross-border expenditure to increase (assuming geographical closeness) and prices will tend to even-out. It becomes difficult to maintain highly differentiated taxes on *goods*, as freedom of imports will allow consumers to buy goods across the border. Integration on an EU scale means that *factors* also become increasingly mobile. Thus, a REA member with high taxes on income would risk losing human capital – often the most well-educated and mobile – to other member states, i.e. a 'brain drain'. Similarly, lower taxes on labour, profits and capital gains will attract investment and can ultimately force a country to lower taxes in order to avoid large scale loss of technological expertise and employment opportunities.

Social policies: The loss of domestic sovereignty in the above can have serious consequences for the social policies of a country. A decline in investment, production, consumption and employment due to factors shifting to other countries basically means that the domestic tax base is 'exported' to a certain extent. The loss of tax revenue and concomitant lowering of social spending due to downward adjustment of tax levels is only part of the issue. The other part is of course that taxes on alcohol, cigarettes and pollution commonly are applied to limit negative externalities, and lower taxes (and increased availability of 'bads' such as alcohol and tobacco) will not sit well with countries which have used taxes and government monopolies as a method of limiting consumption in order to increase societal welfare.[7] I order my cigars via the internet from Spain – which has the lowest taxes on tobacco in the EU – and there is little the Swedish government can do about it. No wonder Spain's total tax revenue from cigarette tax – as a percentage of total tax receipts – is three times that of Sweden's![8]

---

5   A number of economists point to the **internal dynamics** of economic integration, where an initial degree of integration leads to a next step; a FTA will result in cooperation in tariffs, re-export duties and VAT. A base is thus created for additional integration such as harmonisation of standards and policies governing factor mobility between members. The step towards political and not just economic integration is therefore rather short.

6   *Sovereignty, economic integration, and the world trade organisation,* Susan Hainsworth, *D.* Jur. Candidate York University 1996, in Osgoode Hall Law Journal [VOL. 33 NO. 3], page 588.

7   Just in the news; the EU commission is reportedly going to propose that the age limit on purchasing fireworks should be lowered to 12 for all EU countries. This has created quite a stir in several EU countries which have very strict age and purchasing limits on such goods. The Swedes – inventors of automobile safety belts and bicycle helmets – are spitting blood.

8   *Economics of tobacco for the EU region,* Regional Report (EU), June 2001

TRADE

Benefits and costs of a common market

There are many benefits of a common market:

- **Less market distortion** as tariffs are removed and therefore increased **consumer welfare**;

- Increased **competition** and improved **resource allocation**;

- Freer flow of factors will **spread technology** and ideas;

- Increased trade and the possibility of **scale economies** arising from larger markets;

- Increased **labour mobility**;

- And finally, there is a possible '**peace and stability dividend**' arising from diverse cultures becoming closer and interdependent.

There are, of course, several disadvantages:

- Increased harmonisation inevitably leads to far deeper political cooperation where centralised policy-making leads to a **loss of domestic political and economic sovereignty** (see above);

- A larger market and free movement of capital creates a foundation for mergers between firms and **possible monopoly and oligopoly** markets;

- The free flow of goods and the harmonisation of certain taxes – say on alcohol and tobacco – makes it very difficult for a country to run purely **domestic social policies** designed to lower the social costs of such goods;

- Countries with strict **environmental rules and labour legislation** might see standards lowered due to the harmonisation of rules and regulations.

Economic Integration

1. Why might there be both winners and losers due to globalisation?

2. What are the possible negative effects for a country joining a free trade area?

3. What other regulations are necessary for countries entering into a free trade area?

4. What is the main difference between a free trade area and a customs union?

5. Explain why a common market involves more political and economic cooperation than a customs union.

Monetary union (European Monetary Union - EMU)

> 'I know very well that the stability and growth pact is stupid.' - Romano Prodi, EU Commission President, 1999-2004

The Treaty of Maastricht in 1992 clearly stated the EU's movement towards monetary integration; the **European Economic and Monetary Union, EMU**. In 1999 the EURO technically came into being and during 2001 EU currencies (from all 15 EU members except UK, Denmark, Sweden and Greece[9]) were 'locked' in place in exchange rate terms. A European Central Bank (ECB) took over the task of setting and implementing monetary policies for all EMU countries. On the 1st of January 2002 EURO notes and coins were issued. This project created one of the largest shifts in economic power in history. By 2011 there were 17 country members of the EMU.[10]

TRADE

---

9 Greece was allowed to join in 2001. I was there – great party! They went into severe hangover mode around 2008 when they had to pay the bill for the party.

10 Members as of January 2009 were: Germany, France, Italy, Spain, Netherlands, Belgium, Austria, Finland, Greece, Ireland, Portugal, Slovenia, Slovakia, Malta, Luxembourg, Estonia and Cyprus.

531

Monetary policy in EMU

Having a single currency means that two EMU countries cannot have different rates of interest. Just as a single country will have a single monetary policy and thus one interest rate for all regions in the country, a common currency area must have the same system. For example, consider that Belgium had a 4% interest rate while the Netherlands had 5%; savers would simply put their money into Dutch banks and borrow from Belgian banks – and if the rate differential were wider it might even be possible to borrow at a lower rate in one country and deposit it for a higher rate in another country! This means that individual interest rates in EMU countries are set by a common bank, the European Central Bank, and *member countries forego domestic monetary policy*.

Fiscal policy rules for EMU members

In order to keep EMU members roughly in line on the business cycle, key economic indicators must move in relative concert. To enable this – or 'force' it, according to opponents of the EMU – a set of rules were laid down in a **Stability and Growth Pact**, (SGP). The main element in the pact is that members must keep budget deficits within an upper limit of 3% of GDP and overall debt under 60% of GDP. To keep countries in line, penalties of up to 0.5% of GDP can be levied on countries exceeding this limit. This can severely limit a member country's ability to fund expansionary fiscal policies.[11]

Advantages of monetary union (EMU)

The overall benefits of a common currency centre on the benefits of economic integration we have looked at in this chapter:

- **Trade creation:** Having a single currency will facilitate increased trade flows amongst member countries. Consumers will also enjoy greater *price transparency*, since a common currency unit enables consumers to compare prices, which will increase trade flows and lead to…

- **…greater competition:** When households have greater choice in substitutes, firms will have to increase *efficiency* and/or improve products.

- **Benefits of scale:** Increased trade and larger markets will enable firms to *lower average costs* and benefit from scale economies.

- **Lower transaction costs:** Firms dealing in foreign currencies would not have *commission costs* or the risks of dealing with exchange rates.

- **Inter-union and non-union investment:** Increased *predictability and reliability* in dealing with same-currency countries might cause members to invest within rather than outside member countries. Non-EMU foreign direct investment (FDI) and portfolio funds might also be attracted to a market based on a currency which is quickly becoming a competitor to the US dollar – in fact, over 50 countries have pegged their currencies to the EURO[12].

Disadvantages of monetary union (EMU)

There are considerable economic costs associated with a 'one size fits all' monetary policy:

- **Loss of domestic monetary policy:** There will naturally be differences in economic activity and inflation in different countries, so-called '**economic asymmetry**'. While the ECB sets interest rates based on average (expected) inflation in EMU countries, there can be large differences between individual countries. For example, while inflation in EMU countries was an average of 1.6% (annualised) in 2004, there was increasing divergence amongst member countries; Germany, Finland and Austria had 1.0% to 1.2% inflation rates while Portugal, Greece and Ireland had inflation rates between 2.9% and 3.5%.[13] Clearly, the former three might wish to employ lower interest rates in order to stimulate growth, while the latter three countries might then suffer from increased inflation. None of the countries have the option of reducing or increasing interest rates, having surrendered monetary policy to the ECB.

- **Restrictions on domestic fiscal policy:** The de-facto limits on EMU members' domestic use of fiscal policy have been likened to a *'fiscal straitjacket'* as there is clearly a boundary to deficit spending during recessions. It did not take long before an EMU member, Portugal in 2002, exceeded the 3% limit laid down by the SGP, and by 2003 it was clear that half the members

---

11 A new and improved version of the SGP has just been negotiated and should come into effect in 2013.

12 EU homepage at http://europa.eu.int/scadplus/leg/en/lvb/ l25063.htm

13 *Eurostat, Euro-indicators*, news release no 39/2004, 17th March 2004 at http://europa.eu.int/comm/eurostat/Public

of EMU would break the 3% rule during 2004 – with Germany, France, Portugal, and the Netherlands as 'repeat offenders'.[14] While the European Commission has issued public warnings to offenders, no penalties have as yet been imposed on any EMU country.

- **Regional unemployment:** A single currency works well when factor mobility is high; a depressed area will have outward movement of labour and capital towards areas with better employment and investment opportunities. In the US, for example, over 50 million Americans moved during the last three decades of the 20th century. Labour mobility in EMU countries seems to be rather low, perhaps because of language, cultural and remaining bureaucratic barriers. Since regions (or countries) have limitations on using reflationary monetary policies the aforementioned asymmetry in the EMU can create wide divergence in unemployment rates. So while the EURO zone average in January 2012 was 10.7% unemployment, Austria had 4.1% while Portugal and Spain had 15.0 and 23.6% respectively.[15]

Summary & revision

1. A **bilateral trade agreement** is when two countries aim to increase trade between themselves by agreeing on lowering tariff and perhaps non-tariff barriers to trade.

2. A **multilateral trade agreement** is between three or more countries – the WTO is a multilateral trade agreement.

3. A **preferential trade agreement** is when countries are given preferential treatment in terms of lower tariffs on certain goods or types of goods.

4. A **free trade agreement** (FTA) is when member countries lower tariffs and quotas within the FTA area but keep their individual trade barriers towards non-members.

5. **Advantages of a FTA** include increased trade and FDI, reduction in tariffs and thus prices, improved welfare for citizens, and dynamic benefits such as increased specialisation and economies of scale.

6. **Disadvantages of FTAs** include diminished access to FTA members' markets for non-members (trade diversion – HL concept in the next chapter), costs of administering re-export taxes and minimum value-added requirements, and possible short run job losses as firms meet increased competition from FTA area firms.

7. A **customs union** extends economic integration by not only eliminating tariffs and quotas between members but also establishing a *common outer tariff wall*.

8. A **common market** is an elongation of a customs union and occurs when goods, services, investment flows and labour can move freely between member countries.

9. **Problems of further integration** arise, such as the need for supranational institutions which legislate and set rules for all members. There is also a necessary degree of harmonisation of tax laws and social policies. Taken together, a member state will give up a degree of sovereignty.

TRADE

---

14 BBC at http://news.bbc.co.uk/2/hi/business/3607933.stm
15 Eurostat, Euro-indicators at http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&language=en&pcode=teilm020&tableSelection=1&plugin=1

10. **Benefits** of a common market:

    a.   Less market distortion and better *resource allocation*

    b.   Increased *competition*, more *investment*, lower *prices*

    c.   Flow of *knowledge*, *technology* and *production methods* increases LRAS

    d.   Increased *labour mobility* can lower unemployment

    e.   Possible *economies of scale*

11. **Disadvantages** of a common market:

    a.   Loss of domestic economic/social *sovereignty*

    b.   Possibility of *monopoly/oligopoly* markets

    c.   Possibility that some members are forced to *lower their standards* on labour, social and environment legislation

12. **Monetary union** (EMU) entails a single currency and thus a central bank that sets monetary policy for all members.

13. **Advantages** of monetary union:

    a.   *Price transparency* leading to greater competition and lower prices

    b.   Increased *efficiency* in firms, lower average costs (economies of scale)

    c.   *Lower transaction costs* for firms and consumers – this is a trade incentive

    d.   *Increased investment* within the union since firms will not have to deal with exchange rate risks and uncertainties

14. **Disadvantages** of monetary union:

    a.   The most obvious and serious problem is that members *give up monetary policy* – the monetary union central bank sets policy for all members

    b.   *Restrictions on fiscal policies* (such as the stability and growth pact in the EMU)

    c.   *Regional unemployment* can arise in regions/countries since the national government cannot freely run monetary (and perhaps even fiscal) expansionary policies

TRADE

75. Trade Creation, Trade Diversion and Economies of Scale

Key concepts: HL extensions

- Trade creation
- Trade diversion
- Economies of scale

"NAFTA will cause a giant sucking sound as jobs go south." Ross Perot, US presidential candidate 1992

There are a number of additional consequences arising due to economic integration. Countries can both gain (*trade creation*) and lose (*trade diversion*) when relative prices are changed due to regional economic integration.

Trade creation

In joining a customs union, the removal of tariffs between members serves to increase the volume of trade between members, thereby creating trade. **Trade creation** takes place when domestic consumers in member countries import more goods from other members as import prices fall due to the removal of tariffs and quotas – production will shift to lower cost producers. Initial positive effects are the increase in consumer welfare resulting from more goods and lower prices, while *long run* (dynamic) effects include enhanced comparative advantage and increasing specialisation – leading to increasing real incomes and increased trade with non-members. In effect there is a possible expansion in world trade due to a RTA.

Definition: 'Trade creation'

When intra-RTA trade increases and production is moved from higher-cost producers outside the RTA to lower-cost producers within the RTA due to the removal of trade barriers there is **trade creation**. Increased incomes resulting from specialisation and benefits of scale can further this by creating increased demand for imports from non-members.

For example, consider that Japan can produce DVD players domestically and supply them at a price of USD120. Assume that Korea can sell them to Japanese consumers for USD100 but that Japan levies a 30% tariff on imported players, bringing the price of imported DVD players to USD130 for Korean players. Consumers in Japan would predominantly buy domestically produced DVD players.

Now, if Japan and Korea were to form a customs union and tariffs were eliminated, imported DVD players from Korea would replace Japanese VCRs costing more than USD100 since this is Korea's marginal cost. Figure 75.1 illustrates how the tariff is removed and that supply from Korea (assumed to be equivalent to a 'world' supply curve) increases from $S_{Korea+tariff}$ to $S_{Korea\ in\ CU}$. Since the price in Japan falls to USD100 there will be an increase in Korean production and Japanese consumption: Japanese imports have increased from 15 million DVD players to 45 million. This re-allocation of resources shows *trade creation*, since some of the total production has moved to more efficient (e.g. lower marginal cost) producers within the union.
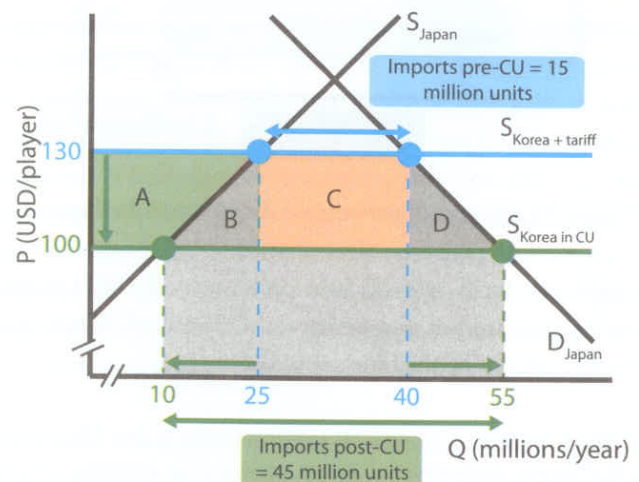
Figure 75.1 Trade creation – DVD players in Japan

535

When Japan enters into a customs union (CU in the diagram) with Korea, imported Korean DVD players fall in price from $130 to USD100 in Japan. Imports from Korea will increase from the initial 15 million units at $S_{Korea+\ tariff}$ to 45 million units at $S_{Korea\ in\ CU}$. The customs union has resulted in trade creation of 30 million additional units from a lower-cost importer.

Other effects:

- A+B+C+D; increase in Japanese consumer surplus
- C; loss of tariff revenue for Japanese government
- B + D; net efficiency gain for both economies

Trade diversion

However, say that at the *initial* tariff of 30% Japan was importing from China rather than Korea since production costs (e.g. marginal costs) in China were USD90. Figure 75.2 shows how Japan would then instead be importing 28 million DVD players from China and selling them for USD117 (USD90 plus 30% tariff) rather than buying these units domestically or importing them from Korea at a market price of USD130 (USD100 plus 30% tariff). The outcome in this case of creating a customs union between Japan and Korea would be quite different from the previous example.
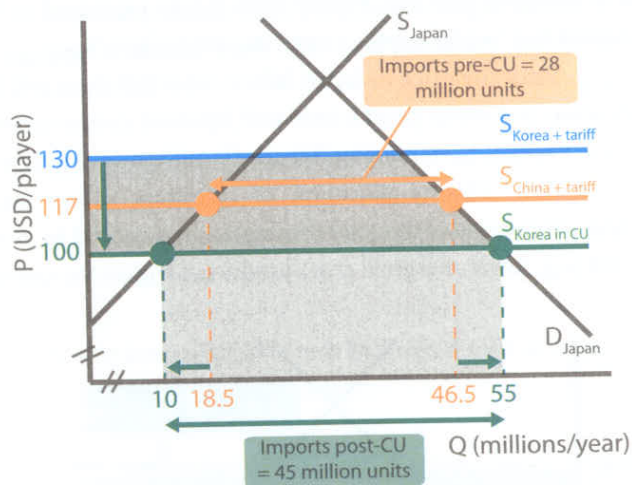
Figure 75.2 Trade diversion – DVD players in Japan

At a Japanese tariff rate of 30% on imported DVD players Korean players are not competitive at USD130 (USD100 plus 30% tariff), shown by the blue supply curve $S_{Korea\ +\ tariff}$.

Japan thus initially imports 28 million units from China at a price of USD117 (Chinese marginal cost of USD90 + 30% tariff). Upon entering into a customs union with Korea, inter-union tariffs on imported DVD players are eliminated, shown

by the shift of $S_{Korea+\ tariff}$ to S_{Korea} in CU and Korean DVD players are now sold in Japan for USD100 which is below the Chinese-plus-tariff price level of USD117.

Japanese consumers increase imports of DVD players from 28 million to 45 million units while the market price falls from USD117 to USD100. Yet since China's exports have been taken over by Korean manufacturers, there has been a **diversion of trade** away from the most cost-efficient supplier (China, at a production cost of $90) to *less efficient* suppliers (Korea, at a production cost of $100) within the customs union.

The tariff-free Korean DVD players at USD100 *divert* trade from the more cost-efficient Chinese players produced for USD90 but price set in Japan at USD117 inclusive of a (discriminatory) 30% tariff. In other words, *lower cost imports from outside the union have been replaced by higher cost imports from within the union*, which is **trade diversion**. In spite of Japanese consumers benefiting from lower DVD player prices and increased consumption, trade flows have been diverted away from the country with a trade advantage, namely China. The example of trade diversion shows that it is not entirely certain that preferential tariff reductions always increase economic efficiency and welfare, since tariffs in the customs union discriminate against non-members. In summa; trade diversion is allocatively suboptimal as resources are diverted away from more efficient suppliers which have a comparative advantage.

Definition: 'Trade diversion'

When a customs union is created and tariff differentials between members and non-members result in trade flows being diverted towards higher cost producers inside the union, **trade diversion** has taken place.

"But," you say, "Japan is indeed a 'winner'! Japanese consumer surplus has increased?! Also, there has been an overall increase in efficiency!" Well, yes ..

... however:

Did Japanese consumer surplus increase as much as it *could* have?! How about efficiency gains? What if China too had been included in the customs union? Doodle yourself a diagram before checking the answer in the Summary and revision

section at the end of the Chapter. In reality, RTAs have *both* trade creation and trade diversion effects. The EU has resulted in considerable trade creation in that intra-EU trade is estimated to be double the level that would be the case without economic integration – where over 50% of the expansion in trade by EU members has been in intra-industry trade. During the first ten years of NAFTA, Mexico's exports grew threefold and per capita income rose by 24% - making it the ninth largest economy in the world.[1] MERCOSUR likewise saw a tripling of intra-FTA trade – within the first four years of inception in 1991.[2] Taken together, numerous statistical studies show that RTAs indeed create trade.

However, all three RTAs above have also resulted in varying degrees of trade diversion. For example, EU agricultural subsidies, tariffs and quotas for sugar block out far less costly production in developing countries, adding some 200% to EU sugar prices. (This diversion of trade to higher-cost producers is augmented by the fact that the EU does not use sugar cane but sugar beets which adds over 50% to the production costs of sugar.)[3] The evidence in NAFTA is strikingly confusing, where studies show a range from 'no trade diversion whatsoever' to 'NAFTA has been primarily trade diverting', yet strong indicative evidence shows that North American imports of textiles from Asia decreased significantly.[4] Perhaps the clearest picture of trade diversion due to preferential tariffs in RTAs is to be found in MERCOSUR, where a preliminary report by the World Bank showed strong evidence that consumers in the FTA were forced into buying more expensive intra-FTA goods rather than cheaper non-member goods which carried high tariffs.[5]

Economies of scale – dynamic gains of customs unions

The effects outlined above are short run or 'one off' effects primarily and thus do not deal with the *dynamic* gains to be had from increased competition and larger scale of firms/industries. In reducing protectionism, member economies will have to shed or re-organise inefficient firms and lower marginal

costs. Recall Smith's quote "…the only limit to the benefits of specialisation are the size of the market…" A customs union means that to all intents and purposes in trade, the border of Sweden (EU member since 1995[6]) is at Gibraltar, the Black Sea and Lake Peipus. Economic theory identifies several key advantages of the economic integration resulting from customs unions:

- Improved *productivity* resulting from increased *specialisation* and allocating production to lowest cost factor markets.

- *Economies of scale* due to increased market size.

- Improved *terms of trade* on the global market – a large customs union can give members 'bargaining power' in competing with firms worldwide.

- Increased *growth* due to the expansion of trade in member countries.

- Increased *spread of production knowledge*, capital and technology within the union.

- Increased *research and development* – larger firms can enjoy abnormal profits which stimulates innovation.

- A large inner market can *attract FDI* from outside the union.

However …

Not everyone is a winner all the time. *Regional unemployment* can increase as resources are allocated abroad and firms outsource to lower-cost labour countries. Increased benefits of scale are partially offset by smaller firms which cannot compete and go bankrupt – leading towards an *oligopoly* structure. Economies outside the union are of course highly disadvantaged – look up 'Fortress Europe + Africa' for example! *Not all goods and markets are equally suitable for scale* economies – many service-orientated economies will not benefit from scale to the same extent that, say, a car manufacturer would.

1   Business Week; *Was NAFTA worth it?*, December 22 2003
2   *Krugman & Obstfeld*, page 244
3   *The Great Sugar Scam*, Briefing paper by Oxfam 2002, pages 2 – 5
4   *An Econometric Analysis of Trade Diversion under NAFTA*, Kyoji Fukao and Toshihiro Okuba, University of Michigan 2002, pages 3 – 5, page 11
5   *Krugman & Obstfeld*, page 245. The authors point out that the original report leaked to the press and created a 'firestorm of protest' from MERCOSUR member governments and the final report was a watered-down version.
6   Yes, agreed, the EU is in fact a common market. It bears pointing out, however, that one of the main 'legs' upon which the EU stands is the *European Customs Union* which has been in place since 1968.

HIGER LEVEL

POP QUIZ 75.1

Trade Creation and Trade Diversion

1. Explain how firms and consumers might benefit from the construction of a customs union.

2. Distinguish between initial and dynamic gains from trade.

3. Assume that, in creating a customs union, member countries agree on a common external tariff that is lower for all goods than before. Might there still be trade diversion?

4. Some economists have termed customs unions 'stumbling blocks for trade' rather than 'building blocks for trade'. Explain.

Summary & revision

1. When a customs union is formed and production is moved to *lower-cost producers* within the union, there is **trade creation**.

2. When a customs union is formed and production is moved to *higher-cost producers* within the union, there is **trade diversion**.

TRADE

3. There are numerous potential **dynamic gains** to be had from a customs union:

   a. *Specialisation*, economies of *scale* and increased *productivity*

   b. Improved *terms of trade*

   c. Economic *growth*, increased *trade* and *FDI*

   d. Increase in R&D due to larger firms and increased profits

4. Potential **disadvantages** of customs unions are:

   a. *Regional unemployment* due to outsourcing and large *oligopoly* firms

   b. All goods and markets are *not equally suited for scale* production

Foregone increase in consumer surplus due to trade diversion.

Figure 75.3 shows that trade diversion does not result in optimal resource allocation. If China had been part of the customs union, the price of imported Chinese DVD players to Japan would have fallen from USD117 to USD90 – which is below the present import price of Korean players at USD100. Japanese imports would be 55 million Chinese units rather than 45 million Korean units. The red area shows the loss of (or foregone) Japanese consumer surplus – and of course this area includes two 'net efficiency-gain triangles'. In summa: trade diversion has resulted in suboptimal efficiency gains.

3.5

76. Terms of Trade

Key concepts: HL extensions

- Definition, measurement and calculation of terms of trade
- Improvement and deterioration of terms of trade
- Short term influences on the terms of trade
- Long run influences on the terms of trade

Definition, measurement and calculation of terms of trade

Assume that Finland is producing mobile phones and trading these for Russian caviar. This is commonly known as the commodity or *barter terms of trade*, as one is using the concept of pure trade without using money.

In adding a *price* dimension, the terms of trade will be expressed as the *ratio* of export prices to import prices. For example, using the example above, one mobile phone costs $US300 and one hectogram (= 0.1 kilo) of caviar costs $US100. The Finnish terms of trade are $US300/$US100, i.e. 3, which means that one mobile phone will buy 3 hectograms of Russian caviar.

The terms of trade for two goods is the price of the exported good over the price of the imported good. As there are thousands of goods being traded, the easiest way to measure the terms of trade for a country is by indexing the values (similar to the method used in creating a consumer price index in the HL extension in Chapter 52). The average prices of export goods and import goods are calculated using weighted indices, whereupon the average price of exports is divided by the average price of imports at a given point in time. For ease of use, the average prices are commonly indexed at a base year value of 100.

Definition: 'Terms of trade index'

The terms of trade index is the (indexed) average price of exports over the (indexed) average price of imports at a given point in time (t_n). A decrease in the terms of trade index shows a deterioration of the terms of trade.

Index of ToT at t_n =

$$\frac{\text{Index of the average price of exports at } t_n}{\text{Index of the average price of imports at } t_n} \times 100$$

Calculate the missing values in the table. Explanation forthcoming under the next heading.

	2007	2008	2009	2010
Av price of X		112	119	123
Av price of M		115	x	125
Index of ToT	100	x	101.7	x

Improvement and deterioration of terms of trade

The base year value of the terms of trade index would of course have a value of 100, e.g. both the average price of exports and imports would be indexed at 100. In 2007 (t_1), the terms of trade have fallen to 97.4 since average import prices have risen by 15% and export prices have increased by 12% ([112/115] × 100 = 97.4). A decrease in the average price of exports *relative* to the average price of imports will render a lower index value and thus a deterioration in the terms of trade. During the next

HIGER LEVEL

TRADE

period import prices must have fallen relative to export prices since the terms of trade have improved.[1] The index in Figure 76.1 shows how the terms of trade for China have deteriorated by 20% over a 10 year period.
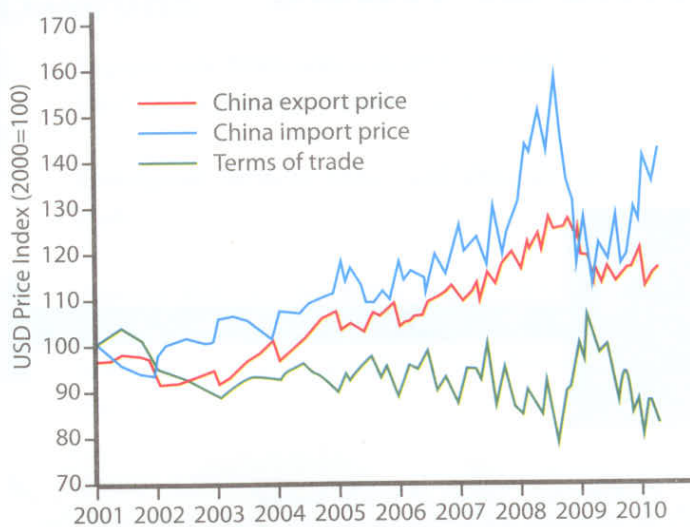
Figure 76.1 China's declining terms of trade over a decade
(Source: World Bank at www.worldbank.org)

However …

Do not confuse 'improved' with better or 'deteriorated' with worse! An improvement in the terms of trade means that the Home economy gets more imports for exports – but it also means that trade partners get *less* for their exports. In the same way, a deterioration in the terms of trade means that Home Country exports are actually cheaper – this might lead to an improvement in current account in the balance of payments.

A 'terms of trade gauge'

The main causes of a change in terms of trade are price changes on the market for goods, changes in the exchange rate and changes in variables linked to AS-AD. (Refer back to Chapter 71 and the 'balance of payments pentagon' in Figure 71.2.) Let us look at two countries, Namibia and Denmark, and two goods produced in each country; salt and B&O headphones respectively. Assuming they trade only theses goods with each other…

Supply and demand and terms of trade: Any change in the price of one good – salt or earphones – will mean that the price of imports for the other country changes, assuming *ceteris paribus*:

- **Demand for B&O earphones:** Say that B&O develop a new revolutionary headphone for the very popular iPods. The increase in demand for the B&O earphones would drive up the price and mean that Denmark will have to export fewer earphones in order to import any given amount of salt from Namibia. *Denmark's terms of trade have improved* while Namibia's terms of trade have worsened.

- **Supply of salt in Namibia:** If Namibia develops improved methods of extracting salt from seawater then the supply of salt will increase driving down the price. More salt will have to be exported to import any given quantity of B&O earphones. The Namibian terms of trade have worsened.

Exchange rates and the terms of trade: Naturally the Namibians don't buy Danish earphones using the Danish crown but the Namibian dollar. Any change in the exchange rate will mean that there will be a change in the relative price of exports in terms of imports for each country.

- **Exchange rate for the Namibian dollar (NAD):** If the NAD appreciates then the price of Namibia's exports will rise in relation to Danish imports and the Namibian terms of trade improve.

- **Exchange rate for the Danish Crown (DKK):** If the DKK depreciates then the price of Danish exports will fall against the Namibian imports – the Danish terms of trade worsen.

AS-AD, relative inflation and the terms of trade: Both aggregate demand and aggregate supply will have an impact on the terms of trade, since both AD and AS will have an impact on inflation and income.

- **Increase in AD in Denmark:** An increase in AD in Denmark and resulting demand-pull inflation will also drive up the price of goods exported – here, earphones. This would improve Denmark's terms of trade. (Note that we are in fact saying that *relative* inflation in Denmark has risen.)

---

1 Note that it is quite possible for the terms of trade to improve due to both lower export and import prices. If the average price of exports falls to 90 and the average price of imports falls to 85, then the terms of trade are 105.9.

- **Increase in AS in Namibia:** An increase in AS would have a deflationary effect on the Namibian economy and the fall in (relative) inflation would also cause export prices of salt to fall. The Namibian terms of trade worsen.

- **Increase in LRAS:** A main long run variable affecting the terms of trade is the gain in productivity in an economy. If Namibia experiences increased productivity relative to its trade partner Denmark then the price of Namibian exports will fall compared with the price of Danish imports.

Figure 76.2 illustrates a 'terms of trade gauge (= measure)' for Namibia and Denmark. As outlined above, there are ultimately three elements which make up the terms of trade for a country; 1) the exchange rate; 2) the price of exports and; 3) the price of imports. A change in any one variable – *ceteris paribus* – will affect the terms of trade (ToT in the figure).

Initially Namibian salt fetches a price of 1,000 Namibian dollars (NAD) per tonne and Danish B&O headphones cost 900 Danish crowns (DKK). The 'gauges' in figure 76.2 (within the dotted box) show an initial exchange rate of DKK0.9 = NAD1, or DKK1 = NAD 1.11. In other words, *one tonne of salt equals one pair of Danish headphones* at the current exchange rate.
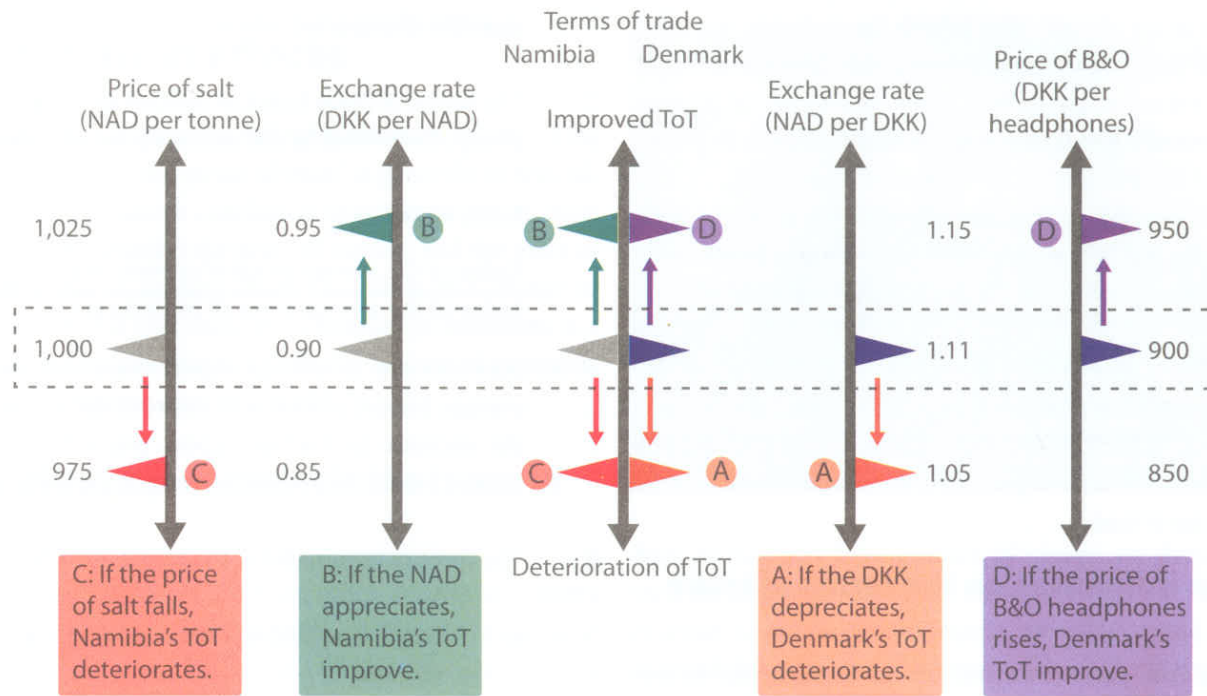
Figure 76.2 Terms of trade 'gauge' for Namibia and Denmark

Now we can play around with the gauges a bit.

- **Depreciation of the Danish Crown:** A depreciation of the Danish Crown (DKK) from 1.11 Namibian dollars (NAD) to 1.05 ('A' in figure 76.2) will cause Denmark's terms of trade to deteriorate, as Denmark will have to sell more headphones in order to get the Namibian dollars necessary to buy a given quantity of salt. Note that in this two-country example, Denmark's deterioration of the terms of trade amounts to an improvement in Namibia's terms of trade.

- **Appreciation of the Namibian Dollar:** Similarly, an appreciation of the Namibian dollar ('B') will enable Namibian exports to be traded for an increased

amount of Danish imports, which is an improvement in the Namibian terms of trade.

- **Fall in the world market price for salt:** Should the market price for salt fall ('C') then Namibia's terms of trade will have deteriorated, since more salt will have to be sold to buy each pair of headphones.

- **Increase in the price of headphones:** This is the same outcome as when the price of Denmark's headphones rises ('D').

HIGHER LEVEL

TRADE

Short term influences on the terms of trade

The terms of trade change over time. In the **short run**, a number of influences will cause a change in the terms of trade:

- Perhaps the most obvious influence on the terms of trade is a change in the **exchange rate** – indeed this is the most likely reason for the decline in China's terms of trade over the past decade as China has strived to keep the Yuan rate low. Depreciation of the home currency worsens the terms of trade while depreciation of trade partners' currencies improves Home's terms of trade.

- Increased **demand for a country's exports** will improve the terms of trade – due to both the exchange rate effect and the price effect on export goods. **Increased supply** of exported goods will lower the terms of trade – especially goods which are demand inelastic in price.

- It is possible to intentionally alter the terms of trade by the use of **trade barriers, intervention purchasing/ selling** of the home currency and **devaluation**.

- A **booming economy** can attract investment funds and cause the currency to appreciate and thus the terms of trade to increase. It is also possible that demand pull **inflation** causes depreciation and a deterioration of the terms of trade.

Long run influences on the terms of trade

In the **long run**, there are a number of fundamental factors which will have an impact on the terms of trade:

- Increased investment and **supply-side policies** which increase long run aggregate supply can lower domestic prices relative to trade partners. Increased productivity would have the same effect.

- Increased world supply and market 'gluts' can depress the market for a country's exports. This happened to a number of coffee and oil producing nations during the 1980s and '90s.

- Increasing incomes will shift demand towards secondary and tertiary goods with higher **income elasticities**, increasing the terms of trade for industrialised countries and decreasing the terms of trade for countries dependent on exports of primary goods.

- **Purchasing power parity** theory (see Chapter 67) predicts that exchange rates will ultimately adjust to different inflation rates in trading countries. Since the terms of trade are strongly linked to exchange rates, PPP adjustment of exchange rates will naturally affect the terms of trade.

Summary & revision

1. The barter terms of trade can be expressed as "'… how many cars do we need to export in order to buy a given quantity of imported wheat."

2. The terms of trade index is defined as the average price of exports divided by the average price of imports.

   Index of ToT at t_n =

   $$\frac{\text{Index of the average price of exports at } t_n}{\text{Index of the average price of imports at } t_n} \times 100$$

3. Any increase in average export prices (or decrease in average import prices) will improve the terms of trade. An decrease in average export prices (or increase in import prices) will of course worsen the terms of trade.

4. The three determinants of the terms of trade are:

   a. Supply and demand for X and M – and thus Px and Pm

   b. Relative inflation rates

   c. Exchange rate changes

5. Long run influences on the terms of trade include:

   a. Supply-side policies and changes in relative efficiency and relative inflation

   b. Gluts or scarcity in world supply

   c. Purchasing power parity and thus the exchange rate

77. Consequences of a Change in the Terms of Trade

Key concepts: HL extensions

- Terms of trade and redistribution effects
- Terms of trade, PED and the balance of payments
- Terms of trade, commodities and LDCs

Redistribution Effects

It is all too common to fall into the trap of regarding 'improvement in the terms of trade' as generally beneficial. This is an erroneous conclusion since the terms of trade is simply a **ratio** between the price of exports and the price of imports and tells us nothing about the effects of a change in the ratio. The fact is that there are a number of outcomes of a change in the terms of trade in both the balance of payments and the macro environment in general.

Effects of an improvement in the terms of trade

Good: Perhaps the most obvious effect is that a country which has improved its terms of trade will be able to consume more imports and thus experience a general increase in *living standards*. Another effect of being able to get more for domestic goods is that *external debt servicing* (i.e. paying off loans and interest) will be easier. Firms will also be able to *import cheaper raw materials* and capital, which can enhance competitiveness.

Improved terms of trade can also improve the **current account** in the balance of payments. If exports are relatively *inelastic*, an improvement in the terms of trade can increase export revenue and improve the current account, since the relative increase in price will be greater than the relative fall in the quantity of export goods sold. The same holds true if imports are demand inelastic; import spending would decrease.

Bad: As you may have guessed, if exports are demand elastic, then an improvement in the terms of trade will cause export revenue to fall – just as import spending would rise if the demand for imports is relatively elastic. Both would have a *negative effect on the balance of payments*. A decrease in export revenue and/or an increase in import spending could lower national income and adversely affect employment.

Effects of a deterioration of the terms of trade

Good: A decrease in the price of exports relative to the price of imports will lead to an *improvement in current account* if the price elasticity of demand for exports is *elastic*. If the demand for exports is also price elastic then export revenue will increase and also improve the current account balance. Increased demand for exports and/or decreased demand for imports will increase *aggregate demand* and perhaps increase job opportunities.

Bad: Higher prices of foreign goods will not only *lower consumption* possibilities for households, but *increase foreign*

debt burdens and make imported factors dearer. A deterioration of the terms of trade will have a negative effect on the current account if the demand for export goods is price inelastic, as total export revenues will fall. Inelastic demand for imports will also be negative for the current account, as total import spending will rise.

Ugly: See terms of trade for developing countries further on.

Terms of trade, PED and the balance of payments

We have dealt with the issue of elasticities in conjunction with exports and imports in some depth already, more specifically in Chapter 73. Let me summarise with a few examples of how different price elasticities of demand for traded goods can affect the **price**, **terms of trade** and the **balance of payments**.

Price elasticity of demand and price fluctuations

As shown in Chapter 9 (*Applied economics; commodity price fluctuations*), primary good prices will fluctuate a great deal more than prices for secondary goods.
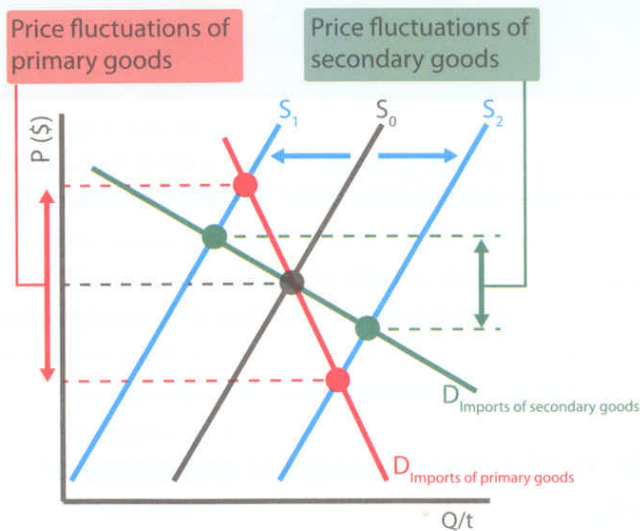
Figure 77.1 PED – price fluctuations for primary and secondary goods

Figure 77.1 illustrates how changes in supply will cause different price fluctuations for primary and secondary goods. Demand for primary goods is more inelastic than demand for secondary goods, as primary goods have fewer substitutes. A disruption, 'shock', in supply such as a poor harvest/flooding (primary goods) or disruptions in the supply of factor inputs (secondary goods) shifts the supply curve from S_0 to S_2. A good season or improved technology shifts supply from S_0 to S_1. The outcome

of decreased or increased supply is of course fluctuations in price, but far more for primary goods than for secondary goods.

PED, terms of trade and balance of payments

The commodity price index diagram further on (figure 77.3) shows that the price of primary goods has fallen fairly consistently during the past 100 years – the exception being the past decade. Diagrams I and II in Figure 77.2 illustrate how the terms of trade for commodity exporters have deteriorated over time; increased supply of both primary and secondary goods has lowered prices over time, but due to low price inelasticity of demand, the price of primary goods has fallen from 100 to 60 – more than that of secondary goods which have fallen from 100 to 80. As the *relative* price of primary goods has fallen more than that of secondary goods, **primary good exporters have had a worsening in their terms of trade**.
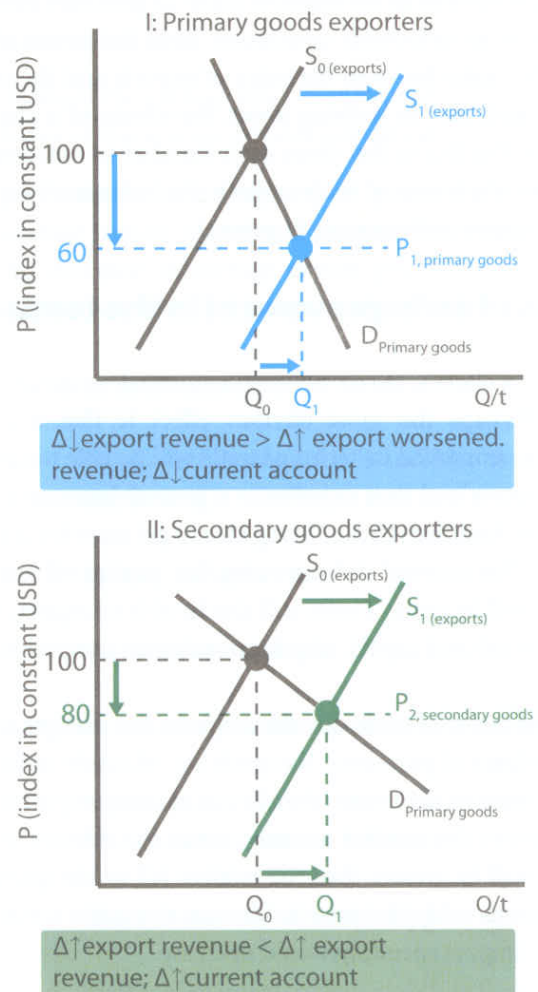
Figure 77.2 PED and terms of trade

The price of primary goods falls relative to secondary good; the terms of trade for primary goods exporters has worsened.

The price elasticity of demand mechanism illustrated for imports in diagrams I and II in Figure 77.2 has further consequences for countries dependent on a narrow range of primary goods for export earnings, namely the negative impact on the *balance of payments*. Diagram I shows how an increase in the supply of primary goods exports **lowers** net export revenue (the revenue gain is less than the revenue loss) while the same increase in supply for secondary goods **increases** net export revenue. This will have a **negative effect on the current account for primary goods exporters** but a positive effect on the current account for secondary goods exports.

Low yED and falling primary goods prices

Adding into the complex equation of PED, falling commodity prices and balance of payments issues is the issue of low income elasticity of demand (yED) for primary goods. Here is how low yED for primary goods causes falling prices for commodities in the long run:

- Low *income elasticity* of demand for primary goods in more developed nations (MDCs) means that even though MDC incomes increase, demand for primary goods increases proportionately less. This is shown in Figure 77.3 as the increase in supply outstripping the increase in demand over the long term.

- Demand has also increased at a slower rate than supply since importers – primarily MDCs – have increasingly found ways to substitute many primary goods. Also, increased efficiency in use of raw materials together with recycling has helped lower the rate at which demand increases over time.

- Improvements in farming methods, new technology in extracting minerals, new methods of locating mineral sources…etc, have all contributed to a remarkable increase in the supply of primary goods over the past 50 years.

> As the increase in supply of primary goods (excluding oil) has been outstripping demand for some 50 years, the long run price trend has been decidedly downwards over the long term.
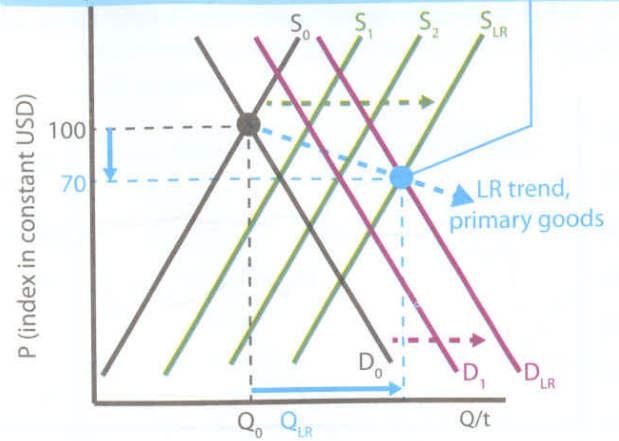
Figure 77.3 Long run supply and demand for commodities and yED

Terms of trade, commodities and LDCs

Developing countries are frequently highly dependent on a few commodities (primary goods) for export earnings. In fact, 62 out of 141 developing countries depended on non-oil commodities for over 50% of all export earnings in 2000 – and if oil is included, the number rises to 95.[1] Increasing supply and in many cases falling demand for commodities has resulted in continuously falling commodity prices over much of the past 50 years. Some commodity categories have in fact fallen by more than 50% in real terms and even the upswing in primary goods prices during the 2000s has not really had an impact on what Figure 77.4 shows to be a long term downward trend in prices for commodities.

> A declining long-term exponential trend of real prices over time
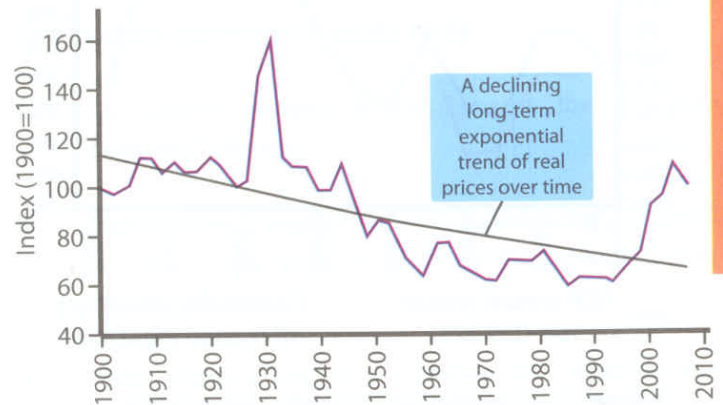
Figure 77.4 Commodity prices over 50 years
(Source: UNDP, "Commodity Dependence and International Commodity Prices" 2010, pages 76 – 79)

1   UNDP, *Commodity Dependence and International Commodity Prices* 2010, pages 76 – 79

Falling commodity prices have severely worsened the terms under which developing nations trade. Figure 4.8.3 shows how the terms of trade (indexed at base year 2000) for developing nations which primarily export commodities have fallen since the 1970s.
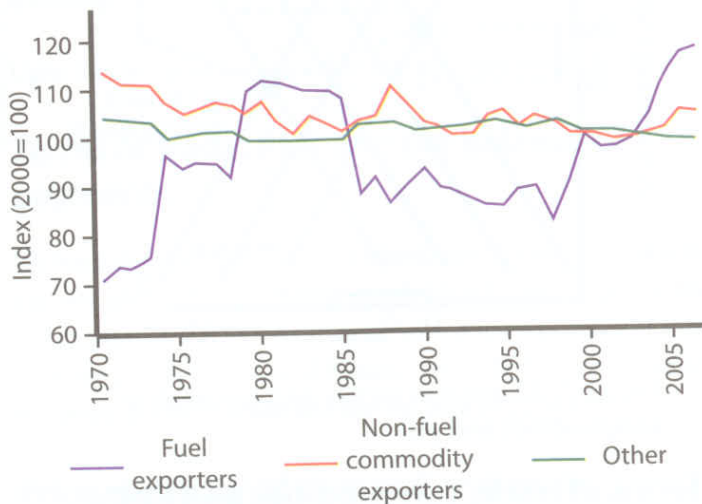
Figure 77.5 Terms of trade for commodity exporters
(Source: IMF Working Paper WP/09/05, *"Commodity terms of trade: the history of booms and busts"*, September 2009)

So how dependent are developing countries on primary goods prices? No words necessary; just have a look at Figure 77.6. Have a good think about this.
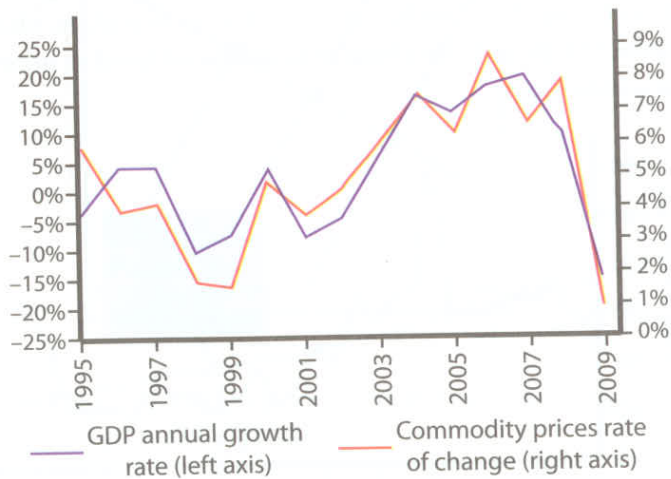
Figure 77.6 Correlation between commodity prices and growth in LDCs (1995 – 2009)
(Source: UNDP, *"Commodity Dependence and International Commodity Prices"* 2010, pages 76 – 79)

POP QUIZ 77.1

Terms of Trade

1. How would a depreciation of the Canadian dollar affect USA's and Mexico's terms of trade?

2. How might a deterioration of a country's terms of trade lead to inflation?

3. Why might an improvement of the terms of trade have negative effects for an economy?

4. How have deteriorating terms of trade affected the majority of less developed countries during the past 40 years?

Summary of terms of trade and IDCS

Let us summarise a key point on deterioration of the terms of trade of developing countries in conjunction with the current account. Falling commodity prices lead to a decrease in the terms of trade for LDCs. As the price elasticity of demand for commodities is low there is a decrease in export revenue and a worsening of the current account.

In economic shorthand: supply for commodities has been outstripping demand for over 60 years $\rightarrow \Delta\downarrow$price of commodities $\rightarrow \Delta\downarrow$terms of trade (**low PED** for exports) $\rightarrow \Delta\downarrow$export revenue $\rightarrow \Delta\downarrow$current account and $\downarrow\Delta Y$.

The results of falling terms of trade for developing countries are almost always negative:

- Higher costs of **debt servicing** as a greater quantity of exports are necessary to earn a given amount of foreign currency with which to repay foreign debt.

- **Current account deficits** often lead to increased borrowing – which in turn increases the debt burden.

- Falling commodity prices often encourage producers in developing countries to **increase production of commodities** – which further depresses the world price of the commodity.

- Deteriorating terms of trade **reduce much-needed imports** such as capital, intermediate products in

production, and fuel. All are needed to industrialise and increase value-added output.

Less developed countries – which have a comparative advantage in primary production – will in all likelihood see their terms of trade decline even further in the near future due to increased capacity and low price elasticity of demand for primary goods.

However...

Australia, one of the largest exporters of commodities in the world, has seen GDP growth of an average 5.3% annually in the latter half of the 2000s – more than half of this growth came from commodity exports. Most of this is due to China's incredible appetite for raw materials. For example, between 2000 and 2010, Chinese imports (in USD terms) of iron increased by a factor of 42.5 and coal by a factor of 248![2] Exports of minerals and agricultural goods made up 57% of Australia's exports in 2012 and exports made up 20% of Australian GDP – which means that 11.45% of export revenue came from primary goods. Figure 77.7 shows the effect on:

- **The terms of trade**: The terms of trade for Australia improved monumentally from 2000 onward – by over 70% just before the economic crises hit in 2008. This was caused by high demand for Australian commodity exports and...

- **...the exchange rate**, which was export-led primarily. As the mining sector expanded to feed China's voracious appetite for raw materials, FDI poured in as foreign mining syndicates established in Australia. By 2010, 32% of *all* FDI inflows to Australia were in the mining sector.[3] This helps explain the current account deficit – billions were coming in on the financial account!

- **Current account**: So, high growth, massive FDI inflows and a strong exchange rate. This indicates a hefty current account deficit according to economic theory. I checked; Australia has something of a current account record in the OECD; the country has been

---

2   http://www.reuters.com/article/2011/11/30/us-australia-china-idUSTRE7AT2HY20111130 and http://www.theaustralian.com.au/business/mining-energy/china-forecast-suggests-commodities-boom-has-peaked/story-e6frg9df-1226291178832

3   http://www.business.nsw.gov.au/invest-in-nsw/about-nsw/trade-and-investment/foreign-direct-investment-in-australia-by-industry

running a current account deficit since 1972. So, while this makes the deficits during the 2000s somewhat less intimidating, one can still see that rising terms of trade and a strengthened Australian dollar almost consistently worsened the current account during the 2002 to 2008.
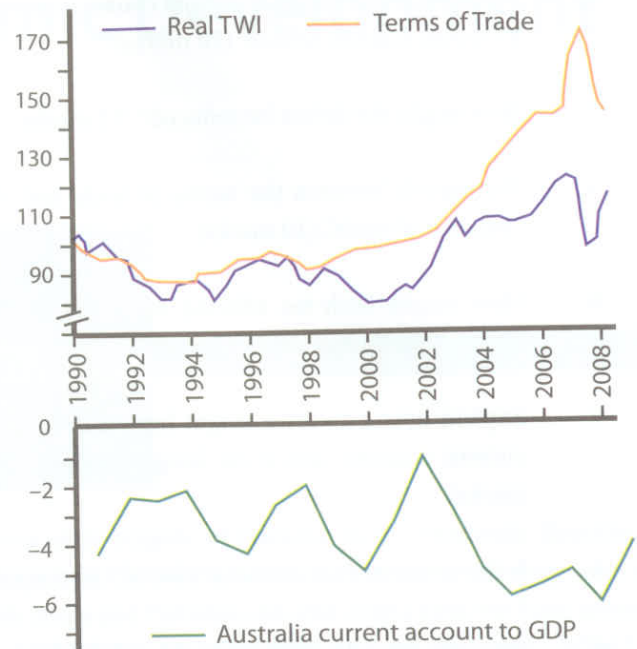
Figure 77.7 Australian current account as a percentage of GDP

(Sources: Reserve Bank of Australia at http://www.rba.gov.au/mkt-operations/foreign-exchg-mkt.html and www.tradingeconomics.com/australia)

Preparing for exams

Short answer questions (10 marks each)

1.   "A current account deficit damages the domestic economy." Discuss.

2.   Explain how a country which is experiencing a "boom" in the domestic economy might see the current account go into a deficit.

3.   How might a country's exchange rate influence the balance of payments?

4.   How might deteriorating terms of trade improve the current account in the balance of payments?

5.   "An appreciation of the exchange rate is always beneficial to an economy". Discuss.

6. Why might a devaluation of a country's currency not necessarily improve the current account in the short run?

Extended response questions (25 marks each)

1. a. What problems might arise for a country running a current account deficit? (10 marks)

   b. How might the deficit be reduced? (15 marks)

2. a. Distinguish between the 'terms of trade' and the "balance of trade". (10 marks)

   b. How might both be affected by a fall in the country's exchange rate? (15 marks)

3. a. Explain how a country might have a consistent current account deficit for longer periods. (10 marks)

   b. Is this necessarily a serious problem? (15 marks)

4. a. Account for the difficulties of establishing a common currency amongst 10 countries. (10 marks)

   b. Discuss the costs and benefits of establishing a common currency. (15 marks)

Summary & revision

1. **Improved terms of trade** have **positive** re-distribution effects:

   a. *Living standards improve* as consumers can consumer more imports and have more choice

   b. *External debt decreases* in terms of export revenue

   c. Firms can import *cheaper factors of production* in terms of exports

   d. Better terms of trade can *improve current account* if exports and/or imports are price inelastic

2. **Improved terms of trade** can also come with **disadvantages:**

   a. *Current account might worsen* since exports are dearer and imports are cheaper

3. A **deterioration** of the terms of trade has **possible benefits:**

   a. As relative export prices have fallen, there might be an *improvement in current account*

   b. Increased exports might *increase aggregate demand* and thus income

4. Possible **disadvantages of deterioration** of the terms of trade:

   a. Higher import prices *impact on consumers in terms of choice and living standards*

   b. Export-driven economies might see the *external debt burden* increase

   c. For export goods that are price *inelastic*, a *deterioration of the terms of trade might lead to a worsening* of current account

5. **Low PED, PES** and **yED** for primary goods has disadvantaged primary goods exporting LDCs:

   a. Low PED and PES has led to extreme *price volatility* over many years

   b. *Low yED for primary goods* has often meant that demand from MDCs for primary goods has not increased in tune with growth

   c. *Supply outstripping demand* has lowered commodity prices in real terms over almost 50 years

6. **Falling commodity prices** have led to a deterioration of the terms of trade for developing nations over most of the past 50 years. Effects of this include:

   a. *Current account deficits* as export revenues fall

   b. *Increased debt burdens* since much of the debt is accrued from the foreign sector and export revenues are needed to service the debt

   c. *Imports of much-needed capital goods become dearer* in terms of export revenues

4.1

78. Economic Growth and Economic Development

Key concepts:

- Definitions of economic growth and development
- A 'virtuous cycle' of development
- A 'spectrum' of development
- Main sources of growth in LDCs
  - ○ Physical capital
  - ○ Human capital
  - ○ Functioning institutions
- Relationship between growth and development

"How does it feel to be amongst the 5% richest humans ever to walk the face of the earth?"

This the question I ultimately ask all my people. My intention is not to create guilt or bad conscience but simply to make them aware of the monumental gulf between haves and have-nots; ability and disability; freedom and confinement; choice and lack of options; and wealthy and poor. I grew up in most privileged circumstances in a number of developing countries in the Caribbean and South America. I was surrounded by servants, diplomats and expatriates at home – and kids on the street who lived in another universe. I knew that there was a gulf separating us but I never really understood why. I still don't.

The answer might lie in education – but how then do you explain the situation of Russia where mathematics is a spectator sport but the population is shrinking and life expectancy falling? Or it might lie in natural resource endowment – but then you look at the utter tragedy of resource rich Democratic Republic of Congo over the past 30 years. How about technology – oh yes, both India and Pakistan can split atoms now, but have amongst the highest proportions of child poverty in the world. What I am trying to get across is that development issues have been studied for some 70 years as a specific field and, as yet, no absolute or generally applicable solutions have arisen. Thus, while this section looks at the problems facing developing countries and some of the possible solutions, there are no patented 'keys' or '12-step programs' available.

This section builds to a very large extent on the core economic issues you have studied in micro, macro and trade. Students in IB economics are therefore rather lucky in having so much revision built-in, yet I would urge you to revise elasticities, production possibility frontiers, limitations of GDP figures and barriers to trade. Development economics is something of a trap for many students, who all too frequently fall into a habit of using normative thinking and loosely linked anecdotes rather than clear economic terminology and theory. I have seen good students fall several grade points on tests in development economics simply because they used sweeping normative statements in their analyses which were not backed up by mainstream economic theory and terminology. Development issues are not 'last' by order of importance, but by order of application: the preceding sections are the **basis** for being able to study development issues.

DEVELOPMENT

Definitions of economic growth and development

Economic growth has been outlined in Chapter 37; it is the increase in real GDP or GNP over a time period. Economic growth is a quantitative variable and lends itself rather easily to comparisons both between countries and over time.

Usage of the term 'economic development' has undergone something of a revolution in the past 40 years. It originally referred to the process by which an economy would shift from a primarily agricultural output to one with a larger proportion of secondary goods, i.e. a process of industrialisation. In this context, the development process was almost entirely an economic reform process, encompassing GDP/GNP increases which outstrip both inflation and population increases – resulting in real per capita growth. One can say that growth has been used as a narrow definition of development.

During the latter part of the 1960s, the view of development in purely economic terms was increasingly seen as insufficient. The failure in many developing nations to show how increasing income 'trickled down' to broader layers of society indicated that simply measuring growth – however 'real per capita' it was – in fact was highly inadequate for showing true development in terms of an overall increase in the standard of living and quality of life for citizens. Developing countries which showed that unchanging portions of the populace were not partaking of income increases and enjoying greater economic equality and employment opportunities could not be considered to have developed. The concept of economic development was therefore broadened to portray not only economic growth, but also a reduction in *poverty*, income *inequality* and *unemployment*.

The United Nations Development Program (UNDP) in its annual Human Development Report, puts it:

> Human development is the expansion of people's freedoms to live long, healthy and creative lives; to advance other goals they have reason to value; and to engage actively in shaping development equitably and sustainably on a shared planet. People are both the beneficiaries and the drivers of human development, as individuals and in groups.

> **Definition: 'Economic growth and economic development'**
>
> **Economic growth** is the increase in real national income during a time period, usually one year.
> **Economic development** is a wider concept, adding to the above definition the overall aims of the reduction of poverty, income inequality and unemployment. Growth is a quantitative concept while development is a qualitative (and thus normative) concept.

A 'virtuous cycle' of development

A good many Catch 22s arise in development issues as shall be seen. Figure 78.1 illustrates how various pre-conditions for human development are linked to forces which enable economic grow; a *pro-development cycle* of growth and development. As looked at in Section 5.1 there are a number of sources which enable economic growth, such as investment in human capital and infrastructure. This in turn will enhance productivity and growth – which then creates improved *conditions* for development.

Means of development: GROWTH

Sources of growth/development:
- Human factors
  - population
  - education system
  - health care
- Institutional factors
  - financial system
  - infrastructure
  - political stability
- Physical factors
  - investment
  - resources
- Other...

Goal: Human development

"Live long and prosper!"

Sources of growth/development:
- Human factors
  - population
  - education system
  - health care
- Institutional factors
  - financial system
  - infrastructure
  - political stability
- Physical factors
  - investment
  - resources
- Other...

Figure 78.1 Pro-development cycle

DEVELOPMENT

It should be noted that Figure 78.1 is indeed optimistic and perhaps even a trifle misleading. The UNDP has conducted numerous studies on the links between growth and development, and while indeed the links exist, they are not automatic. In other words there is no evidence to support that growth inevitably leads to development, but that increased national income must be 'translated' (UNDPs choice of words) into development by sound policies and well-functioning institutions and infrastructure.[1]

A 'spectrum' of development

So, economic growth is computational by nature, i.e. a matter of defining output and adding up the numbers, while economic development is more complex since it involves looking at several aspects of society rather than simple numeric values of national income. Clearly development is desirable and thus it is normative in nature where the rule is 'the more the better'. One can say that economic development as a concept takes into account the questions of *what* and *who*; what is being produced and who is getting it? Addressing these two questions using figures for poverty levels, unemployment levels and distribution of income provides a basic way of outlining economic development since all three are closely linked to the creation and sharing of national output.

Why growth is not the same as 'development'

The following questions can illustrate the issue: In an economy where real **GDP per capita** has increased without affecting the poorest 25% or 50% of the population at all, has the economy developed? And does economic development take place if, during economic growth, a large proportion of the population is unable to meet basic needs for food, shelter and sanitation? Does economic development take place when long term unemployment increases in a growing economy due to an increase in productivity? If economic growth is fuelled by a few wealthy land owners producing primary goods for the foreign export market rather than for the domestic market, and the profits are re-invested abroad, can one speak of economic development?

Most development economists answer 'no' to all the above. (But I hope you also realise that I am being a bit provocative!) By bringing into economics a *social dimension* of the dispersion of output and income, e.g. value judgements on what is socially desirable, the definition of economic development has undergone a radical change. Economists now incorporate

'income equality' into economic theory and put forward suggestions aimed at alleviating poverty and increasing income distribution. This is done without a great deal of reflection on normative issues involved – yet, as we shall see further on, pure *economic* arguments are indeed put forward in support of development.

Simply put, economic growth which does not enhance the **living standards** of a broader measure of the population is defined as sub-optimal in developmental terms; widening income gaps between rich and poor are bad while narrower gaps are good. This includes effective allocation of labour resources in keeping unemployment low and also a 'fair' system of taxation and social benefits to even-out income gaps. It was seen quite clearly during the 1970s that growth did not measure up to these development criteria since income gains were often not distributed across a wider spectrum of society.

A development 'spectrum'

Figure 78.2 illustrates how the concept of development is successively widened. From a narrow definition of real GDP/GNP per capita, economic development expands the conceptual framework to include people's access to the benefits of growth; employment and a share of the output increase. By adding on a number of socio-economic indicators/measurements, development as a concept went from the measurement of pure economic welfare to what we now attempt to measure by using an array of variables. 'Wider' development (note: my own choice of words) encompasses any number things which enable people to enjoy richer lives:

- **Political freedoms** such as voting rights and equality under a rule of law;

- **Access to education** and a choice in careers;

- **Social freedom** of free time – and having choices of cultural activities to fill this time;

- **Societal independence** in terms of freedom of indebtedness and/or indentured (= contracted) labour service.

Ultimately, a high level of development is primarily **enabling** and **empowering**. It enables personal freedoms and choice and empowers people to make these choices.

1 See *HDR* 1996, Chapter 3 passim and HDR 2003 page 17
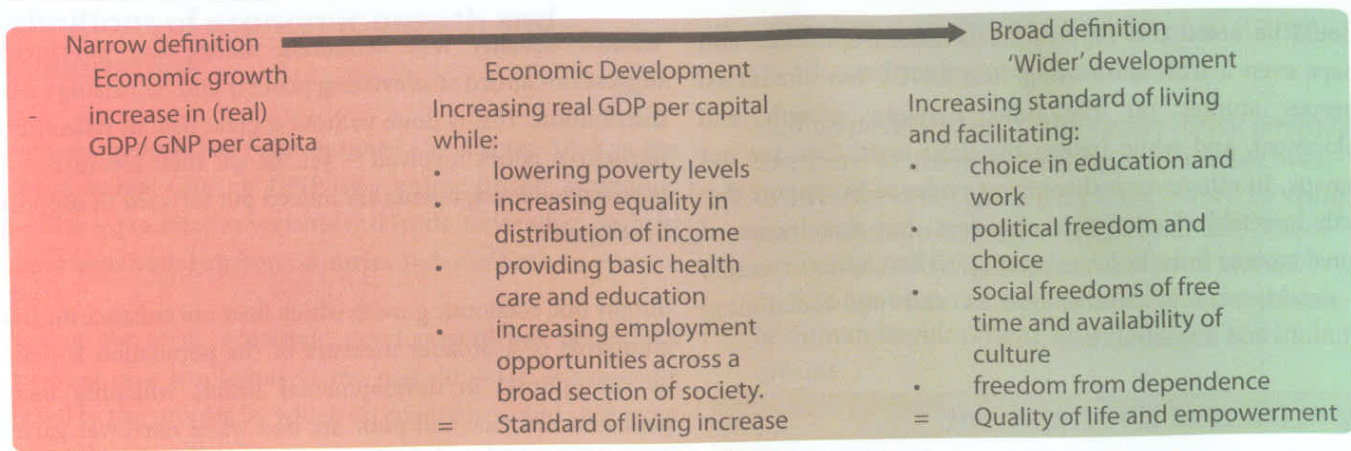
DEVELOPMENT

Figure 78.2 Growth and development spectrum

The progression in Figure 78.2 is my own version but builds strongly upon the work of Nobel Laureate Amartya Sen – who helped develop the Human Development Index (HDI).[2] As I have stated earlier, development is hard to define and the table above is an attempt to start from a basic agreed-upon basis – growth – and widen the definitional span in stages. Whereas pure economic development stresses such elements as the existence of jobs, a viable workforce and a taxable income base, development stresses the wider context of an *enabling function*. It enables people to enjoy the wider benefits of life outside meeting the basic necessities of food and shelter. The 'spectrum' shows how the concept of development has progressed from being equated with economic growth to the far wider definition of development. It should be noted that a number of additional variables shall be included in the 'wider' definition of development in Chapter 82.

Main sources of growth in LDCs

It is pretty clear that while natural resources are vital in both economic growth and development, they in no way explain the whole story. A good many poor countries have abundant natural resources and are still in early development stages. A few brief examples:

- The oil reserves of **Saudi Arabia** are estimated to account for one third of all known reserves, yet the country ranks 56th in the 2011 Human Development Index (HDI – see Chapter 82).

- **Nigeria** has oil, virgin rainforests filled with hardwoods, tin, iron, coal, palm oil trees, cotton, bananas, peanuts (called 'groundnuts' by the British)….and a good deal more. How does Nigeria measure up on the HDI? The country is one of the poorest in the world and had an

HDI value of 0.46 in 2011, placing Nigeria at 156th place of 183 countries included in the index.[3]

- The 'Asian Tigers' Hong Kong, Taiwan, South Korea and Singapore have nowhere near the natural factor endowments of Nigeria but are now amongst the richest countries in the world, ranking well into Very High Human Development in the HDI.

While commodity production can be a vital source of income in the short term in LDCs, it is highly unlikely that primary industries such as mining, agriculture and fishing can fuel development in the long run. Increasingly the focus has been on *physical capital*, *human capital* and *institution building*.

Physical capital

As outlined in Chapter 49, any accumulation of capital will enable increased output in the future – sustained growth in an economy will necessitate increased capital stock, which means that there must be continuous (net) investment taking place. This often leads to a most difficult 'chicken or egg' problem for LDCs, since savings are needed for investment, income is needed to create the basis for saving; and investment is necessary to increase income. So which comes first? See Chapter 55 for a non-answer.

Technological factors: The basic definition of technology as applied to economic development is simply 'method of production', so a man-pulled plough is one form of technology and a tractor another. New technology is any new and improved *technique* in production which increases output per unit of input, and technological progress will increase output and growth over time. New grain types and improvements in irrigation techniques; innovation in solar cells; and mechanisation in garment production are all examples of technological improvements.

2   See pages 5 – 11, *Development as freedom*, and also a very accessible account in *Todaro*, pages 14 – 24.

3   Human Development Report (HDR) 2011, Table 2, page 133

In summing up, economic growth is a **function** of land, labour, capital and efficiency. Any increase in the amount or improvement in quality will enable higher output. Note that investment can come with an assortment of price tags attached. The most obvious is the trade-off between consumption and investment, i.e. the **opportunity cost** of present consumption. This can be felt rather harshly in LDCs which might have little choice but to use resources for present needs – simply not being able to afford to wait. Just do a little thinking on what "…afford to wait…" means until Chapter 95.

Another issue is the **appropriateness** of the technology in question. While increasing availability and use of capital will have many positive effects such as the spread of technological knowledge and capabilities in developing countries, there can be negative side effects due to the labour-saving element in technological progress. Technology which makes a few hundred workers redundant might have far higher societal costs than economic benefits. In addition, advanced machinery needs to be installed, maintained and repaired – all of which often involve imported parts and expertise which will need to be paid for in scarce foreign currencies. It is therefore argued that using abundant relatively cheap labour often results in higher overall social benefits than imported labour-saving capital.

Human capital

If India can turn into a fast-growth economy, it will be the first developing nation that used its brainpower, not natural resources or the raw muscle of factory labour, as the catalyst.[4]

Growth of human resources: The growth/development potential resulting from human factors of production can be narrowly defined as more and/or better labourers. Both serve

to shift the PPF outwards (see Section 1). Population growth adds to overall potential output – whether it is a matter of more people being born or immigration. However, overall economic growth resulting from an increase in population does not necessarily mean that average incomes rise. In fact, a majority of the 40 or so **least** developed countries saw GDP per capita decline in the 25 year period between 1975 and 2000.[5]

Yet this should not detract one from the simple fact that people are a *resource* and that more resources add to potential. Demographic changes can render amazing outcomes in a very short period of time. Ireland was one of the poorest countries in the EU in 1980 but had 50% of the population under 25 years of age. This large group of young people contributed greatly to making Ireland one of the richest in the EU 20 years later. India is in a similar situation, with a large young person to dependant ratio, and a number of economists now talk about a 'window of development opportunity' for India in the next 20 years.

Increased quality in human resources, e.g. human capital, means the increase in knowledge, skills, education and training of the labour force (see Chapter 49). The economic benefits include increased labour mobility, increased incomes and tax revenues, higher productivity and, of course, growth. The social benefits of education include higher life expectancies due to increased income and better health care, more opportunities – particularly for women – and wider overall participation in society's offerings of culture and politics.

Again, the above is a short-list. It is almost impossible to quantify and assess the benefits of human capital investment, but there are a number of strong indications. Figure 78.3 shows the comparative literacy rates, life expectancies and real GDP in purchasing power parity for six countries over a 45 year period.

4   *The Rise Of India*, Business Week, December 8 2003

5   HDR 2002, pages 192 – 193

Figure 78.3 Comparisons in human development, selected countries (2006) Source: Composed from HDI 2008

Country (HDI value/rank)	Life expectancy (years)		Adult literacy (% of population)		Real GDP per capita (PPP$)	
	1960	2006	1960	2006	1960	2006
INDIA (0.609/132)	44	64.1	34	65.2	617	2,489
BOTSWANA (0.664/126)	45.5	48.9	41	82.1	474	12,744
INDONESIA (0.726/109)	41.2	70.1	54	91	490	3,455
CHINA (0.762/94)	47.1	72.7	..	93	723	4,682
THAILAND (0.786/81)	52.3	70	60	93.9	985	7,613
SOUTH KOREA (0.928/25)	53.9	78.2	88	98.5	690	22,985

DEVELOPMENT

This selection of countries in Figure 78.3 does not enable far-reaching conclusions. It is mostly an example of how education comes hand-in-glove with development. Life expectancy and real GDP per capita (PPP USD) were similar in India, China and Indonesia in 1960.[6] Over the next 45 years both China and Indonesia pulled ahead in literacy, life expectancy and GDP per capita. An even greater increase in life expectancy and GDP per capita was recorded by South Korea where per capita income increased 1300% during the period.

While one must realise that the figures do not show whether education and health care *caused* growth or vice versa, the *correspondence* between growth/development and investment in human capital is clear. An in-depth study on this correlation by the World Bank in 1993 suggested that one of the main factors in the strong economic development in East Asia was that a significant proportion of resources went towards basic education.[7]

Functioning institutions

Institutions do not necessarily mean a physical building or place. Instead, the concept is broadly a set of rules and norms which come together under organisations. The institutions in the captions below are often considered the *result* of a developed and well-functioning economy, but increasingly the flip side of the correlation coin is pointed out: these institutions play an integral part in the *process* of development.

Banking system

A banking system has one basic purpose; to *provide liquidity* (money) on the capital and goods markets. The system does so by offering a return to depositors and setting a price for borrowers – interest. In this manner, banks function as a means of putting potential entrepreneurs in contact with the individual savings of others. A bank thus facilitates (= makes possible) finance and the efficient allocation of funds by supplying a system for savings and investment. In acting as a conduit (= channel) for investment, a modern banking system provides a number of vital services in a money-based economy (see Chapter 83).

Education system

One of the highest rates of return in both economic growth and development issues is undoubtedly to be found in education. Every country in what we now call the developed world experienced the brunt of growth and development once primary education had been established. In fact, one of the world's most famous development economists, Nobel Laureate Amartya Sen, points out that Japan as early as 1870 had higher literacy rates than Europe – allowing rapid transition to an industrialised power 50 years later. Sen points out that the 'miracle' economies of East Asia had similar levels of literacy.[8]

Infrastructure

Improved infrastructure lowers transportation times and costs[9] – which can be up to 40% higher in LDCs (as a percentage of GDP) than in MDCs. Good infrastructure thus has a large impact on national income and reduces poverty. World Bank studies show that the increase in investment in infrastructure in LDCs in the 1990s reduced poverty levels by as much as 2% in low-income countries.

Poor water and sanitation facilities result in around 4 billion cases of diarrhoeal diseases every year – causing between 1 and 2 million deaths. **Clean water and basic sewer systems** would reduce deaths and enhance health tremendously – a conservative estimate puts an eightfold increase in benefits to society for every US dollar spent here.[10] Child mortality has been shown to have fallen by 55% owing to access to clean water and the WHO estimates that some 80% of all diseases and a third of all deaths in LDCs are caused by lack of adequate drinking water and sanitation facilities.

Adequate **roads** enable fresh produce to reach markets. Roads allow capital assets to be transported to remote areas which need resources, such as for the building of dams for power plants. Roads also help to get children to school – an especially significant increase in' school attendance by girls has been observed.

There are external effects of adequate infrastructure, namely the ability of LDCs to attract **foreign direct investment**. Firms which are willing and able to invest abroad are primarily interested in keeping costs down, quality of goods up – and making a good rate of return. Poor infrastructure in LDCs

6   In the previous edition I used Botswana as an example and the time span was 1960 t$_0$ 1993. Botswana's life expectancy in 1993 was 65.2 years. By 2006 life expectancy had fallen to 48.9 according to the UN figures used here. (39 years according to USAID.) A decrease in life expectancy of 25% over a period of 13 years?! Look it up and tell me why this has happened.

7   World Bank: *The East Asian Miracle, Economic Growth and Public Policy*, Oxford University Press 1993, passim. I hasten to add that later Human Development Reports from the UNDP find a very weak link between economic growth and significant improvements in health and education for low to middle income countries! See HDR 2010 page 4 for example.

8   *Development as freedom*, Amartya Sen, page 41

9   A reporter for the *Economist* wrote about travelling with a lorry transporting Guinness in Cameroon. The 500 km journey took four days and arrived with one third of the brew missing – bribes to 'robber-cops' at road blocks. (See *Trucking in Cameroon*, Dec 19th 2002.)

10 *The stuff of life*, The Economist, May 15th 2004

helps explain why close to 90% of all FDI takes place between developed countries.

Figure 78.4 ranks a number of countries according to GDP per capita (PPP USD – see Chapter 82) from lowest to highest. The level of infrastructure, measured by 4 variables, is a clear indicator of relative wealth but also implies the potential for countries' future output and growth.

Figure 78.4 Infrastructure and GDP per capita (USD adjusted for PPP)

	Telephone mainlines per 1,000 persons	Paved roads, % of total	% of population with access to safe water, 1995	Irrigated land % of cropland 1994-96
Mozambique	4	19	9	3.4
Tanzania	3	4	49	4.6
India	19	46	85	32.0
Jamaica	140	71	93	14.0
Romania	167	51	62	31.4
Brazil	107	9	69	4.9
South Africa	107	42	59	8.1
Chile	180	14	91	32.6
New Zealand	486	58	90	8.9
Spain	403	99	99	17.7
Singapore	543	97	100	na
United States	644	61	100	12.0

(Source: *WDR* 1997)

Political institutions

"A liberal democracy is a system whereby the rulers get votes from the poor and money from the rich while promising to protect each from the other." *Kevin McGee*, brother and cynic.

Imagine trying to work efficiently in a company where the leadership changed every year; where rules were altered continuously or were simply paper products which had no applicability in reality; where disputes were settled by whom one knew in power rather than by equal rules for all; where people got fired arbitrarily; and where orders, contracts and shipments vis-à-vis other firms were honoured only part of the time.[11] Now put the same board of directors in charge of a country.

People who have the right to elect their rulers commonly become better 'subjects' or simply citizens. Rules and regulations which govern our actions are far easier to abide by if one has been allowed to participate in the legislative process – however indirectly. Functioning **democratic institutions** which result in free and fair electoral processes and which give equal voting rights to all is common to virtually all developed countries.[12] Open and transparent government institutions must also have an external system of checks and balances; freedom of expression, organisation and press. Democracy, as a 'soft' or humanitarian system of rule, has been proven to be strongly pro-growth – perhaps simply because democracy allows and encourages ideas, enterprise and social responsibility. Bad rulers don't survive elections in a democracy and it bears repeating that there has never been a famine in a democratic country.[13]

Most – but far from all – economists would agree that developing nations will need a larger element of planning than developed nations. And while there is debate as to whether a democratic system (India) is more efficient than an authoritarian system (China), economists generally agree on the empirical evidence showing that development is more likely to take place when basic political foundations for the rule of the land are in place.

Relationship between growth and development – a lengthy discussion cut short here

A good deal of the basic outline between growth and development was brought up in Chapter 49. The basic questions on the table for discussion here are:

- Is growth the same as development? No.

- Is there a correlation between growth and development? Yes, but there are some heavy qualifications. Growth means increased incomes and increased general consumption for citizens – keeping in mind that income per capita is an *average* and not all citizens will benefit equally from economic growth.

- Does growth *cause* development? Yes, commonly, but not always and not in all cases. Growth commonly

11 I think I just described a school I've worked for.
12 The jury is still out on Singapore's development.
13 See *Sen*, Chapter 6

555

increases the size and breadth of tax bases (= types of economic activities that are taxable) and provides governments with more funds for public and merit goods. Having said that, there are also numerous counter-developmental issues of growth, such as equity issues, environmental issues and the disruption of traditional societies.

- And finally; can development *cause* growth?! Yes – and this is part of the discussion; which of the two comes first and to what extent is there dual causality. Economics takes the view that increased human capital (increased health and education) results in higher productivity and thus growth.

I have given the standard text-book answers to the four points above – answers which I have also given in class for some 15 years now and 10 years as an economics author. At this point simple honour and academic honesty compel me to admit to two things: 1) It seems that I have been quite wrong about a few issues in the growth-development debate and; 2) I am now more confused than ever. Perhaps this is not quite what a young person who is facing IB exams needs to hear from a teacher/ author but I have long since held honesty and frankness in far higher esteem than 'face' or 'pride'. I do what Keynes did; in discovering I am wrong, I change my mind. What do you do?

Human Development Report 2010[14]

So here it is: there is little – if any – hard-line evidence that there is correlation between economic growth and development in LDCs! I am going to avoid the more complex issue of causality and stick to the more straightforward issue of a pattern or link between the two variables. Figure 49.4 (Chapter 49) showed clearly that economic growth and development are strongly correlated – while Figure 78.5 shows that there is almost no correlation whatsoever!

Figure 78.5 Non-income HDI correlated with economic growth

The answer lies of course in the different Y-axes; Figure 49.4 includes GNI in the HDI while Figure 78.5 here uses only education and health care. The HDI is comprised of three sets of indexes; life expectancy at birth, education level, and GNI per capita in PPP. It turns out that the high value of correlation one gets between income and development measured by the HDI is grossly misleading since the HDI *includes* national income! It's like doing a correlation study where you have 'IB economics performance' on the Y-axis and 'Class attendance' on the X-axis – and 'IB economics performance' is measured by the variables 'Percentage of total marks attained on average' plus 'Percentage of classes attended'. Basically half of the correlation study is measuring the strength of the link between 'attendance' …and 'attendance' – which should be pretty strong, no?

In removing GNI from the HDI in Figure 78.5 we get the correlation between economic growth (X-axis) and development measured solely by *life expectancy* and *education levels* (Y-axis). The line of best fit shows extremely weak positive correlation. A hard-line summary from the HDR 2010 in a few words would be this: "..the correlation between GDP per capita growth and non-income goals [Millennium Development Goals] is practically zero."[15]

The WDR gives the following example of lack of correlation[16]: A young girl born in China (notably one of the fastest growing economies in the world) in 1970 could expect to live 63 years. A young girl born in Tunisia the same year had a life expectancy

---

14 The Human Development Report (HDR) is issued since 1990 by the United Nations Development Program (UNDP). See their excellent site at http://hdr.undp.org where you can have a go at concocting your very own development indicator.

15 WDR 2010, page 47
16 Ibid

DEVELOPMENT

of 55. After 30 years of 8% per annum growth in China and 3% in Tunisia, a girl born in Tunisia today can expect to live 76 years while a girl born in China 75 years.

I hastily point out that the iteration above clearly states "little to zero correlation" between growth and development – not negative correlation! In other words, growth is not anti-developmental and development does not hamper economic growth. But the recent findings of non-correlation between the two seem to indicate that *growth is not a pre-requisite* for development.

This debate isn't over! See www.goodbadeconomist.com

Summary & revision

1. **Economic** growth is a quantitative variable and is measured by the increase in GDP/GNI per capita and adjusted for inflation. **Development** is a broader and qualitative measure of living standards and usually is defined using a composite index such as the HDI.

2. **Development as a concept** commonly includes measures of *education*, *poverty*, *employment* opportunities, *income distribution*, and access to basic *health care*.

3. **Main sources of growth** in LDCs include:

    a.   *Physical* capital

    b.   *Human* capital

    c.   Functioning *institutions*

4. **Growth and development** are commonly thought to be positively correlated – yet recent studies show that this *link is in fact very weak*.

79. Characteristics of Less Developed Countries (LDCs)

Key concepts:

- Common characteristics of LDCs
  - Productivity and income per capita
  - Unemployment and underemployment
  - Agricultural sector
  - Informal economy
  - Income inequality
  - Corruption
  - Poverty levels
  - Birth rates, death rates and population growth
  - Urban sprawl (Todaro model)
  - Poverty trap and pro-development cycle(-s)
- Range and diversity of LDCs
  - History
  - Geography and climate
  - Resource endowment
  - Political systems and political stability

There is a most confusing and even somewhat contradictory array of terms used in development economics to classify countries. One of the standard ways of classifying the world was 'First World' (rich countries), 'Second World' (planned/command economies) and 'Third World' (poor countries). As the Second World no longer exists as an identifiable entity, this classification is pretty much obsolete (= outdated).

Another common division is the 'North – South' division, where rich countries lie in the north and poor in the south. What about Australia, New Zealand and Singapore (OK, Singapore is just in the northern hemisphere)? Then it gets even trickier when one speaks of 'industrialised' vs. 'industrialising' countries; Russia is most definitely an industrialised country but has seen average income fall drastically – around 30% – between 1990 and 2001 and had an average growth rate of -0.1 percent between 1990 and 2005.[1] In addition, you will often see references to '**Newly industrialised countries**', NICs, which is perhaps even more wide of the mark as the term often includes Korea, Thailand and Hong Kong – which have been rather industrialised since the 1960s.

The distinctions are academic rather than useful. I will limit myself to using 'developing country' and 'developed country' or 'less developed country' (LDC) and 'more developed country' (MDC).[2] In some specific cases, reference will be made to high, low and middle income countries (see table in Figure 79.1) in accordance with the World Bank classifications.[3]

---

1 HDR 2003, page 53, and HDR 2007/2008 at http://hdrstats.undp.org/indicators/135.html

2 One can normatively say that a developing country is one which has several of the following characteristics; poor health care; low productivity and GDP per capita; poor standard of living; lack of capital; and a dependency on agricultural produce and raw materials. In this characterisation, circa 148 of the world's countries are developing countries. It is worthy of notice that there will be very large differences between these countries, from dictatorships to democracies and from countries suffering recurring famines to countries which are industrialised export nations.

3 See http://data.worldbank.org/about/country-classifications/country-and-lending-groups#High_income

Figure 79.1 UNDP classification of countries (2012)

High income countries (GNI/ capita over USD12,276 in 2012)	Upper-middle income countries (GNI/ capita of USD3,976 to 12,275 in 2012)	Lower-middle income countries (GNI/ capita of USD1,006 to 3,975 in 2012)	Low income countries (GNI/ capita of USD1,005 or less in 2012)
70 countries	54 countries	56 countries	35 countries
Australia	Algeria	Angola	Afghanistan
Canada	Brazil	Bolivia	Cambodia
Denmark	Mexico	Egypt	Haiti
Ireland	Malaysia	Indonesia	Dem. Rep of Korea
Japan	Russian Fed.	Sudan	Myanmar (Burma)
USA	Thailand	Ukraine	Zimbabwe

Common characteristics of LDCs

While it is worthwhile to warn against attributing too wide a range of common traits to LDCs, there are a number of common denominators.

Productivity and income per capita

Productivity (dollar output per unit of input – commonly labour) is considerably lower in developing countries than in more developed countries. This has mainly to do with a higher capital to labour ratio in MDCs and of course considerably higher technological and human capital levels. Figure 79.2 illustrates that both low income countries (Senegal and Ethiopia) and upper-middle income countries (Mexico and Brazil) have considerably lower productivity than high income countries such as Germany and Japan.
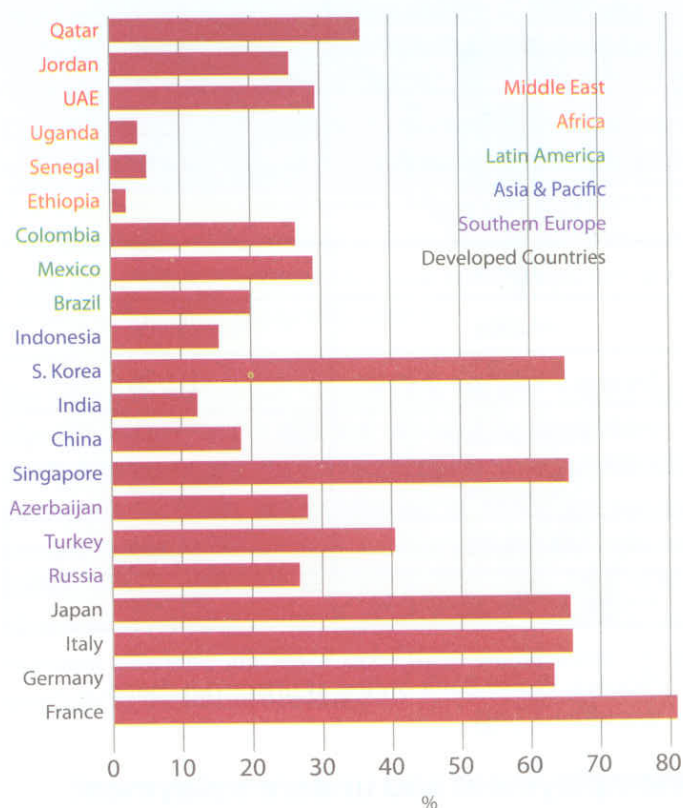
Figure 79.2 Productivity (2010) in an assortment of countries as a percentage of US productivity

(Source: International Labour Organisation at http://kilm.ilo.org/manuscript/kilm17.asp)

Low national incomes across the board can be observed in developing countries. Figure 79.3 shows the same UNDP groupings as earlier but for 2010.[4] I have intentionally selected slightly different figures than those used in Figure 79.1 earlier – I have used GNI adjusted for purchasing power parity (PPP, see Chapter 81) to show that in terms of actual spending power, most LDCs will have somewhat higher PPP incomes than when shown at a given exchange rate. The high income countries' average income is more than 28 times that of the low income nations and more than nine times the lower-middle income countries.

Figure 79.3 GNI per capita in PPP (2010) for income groups (UNDP classification)

UNDP classification	GNI per capita in PPP (2010)
Low income	
Uganda	1,250
Ethiopia	1,040
Average for the group (35)	1,190
Lower-middle	
Indonesia	4,200
Sudan	2,030
Average for the group (56)	3,730
Upper-middle	
Algeria	8,100
Mexico	14,400
Average for the group (54)	11,730
High income	
Australia	36,910
Denmark	41,100
Average for the group (70)	34,061

(Source: World Bank online at http://databank.worldbank.org)

Unemployment and underemployment

High unemployment rates in LDCs are often 'hidden' due to weak government agencies not reporting actual figures and that many people are highly *underemployed* – e.g. are working far

[4] The latest year where comparable figures are available for most of the countries.

fewer hours than they would be willing and able to. There are blank spaces in the World Bank and ILO data banks you can park a battleship in – there is simply no data for many countries. Figure 79.4 should thus be taken with a large grain of salt.

- North Africa
- Central and S.E. Europe (non-EU) & Central Asia
- Middle East
- Developed Countries
- Sub-Saharan Africa
- Latin America & the Caribbean
- South Asia
- East Asia & the Pacific
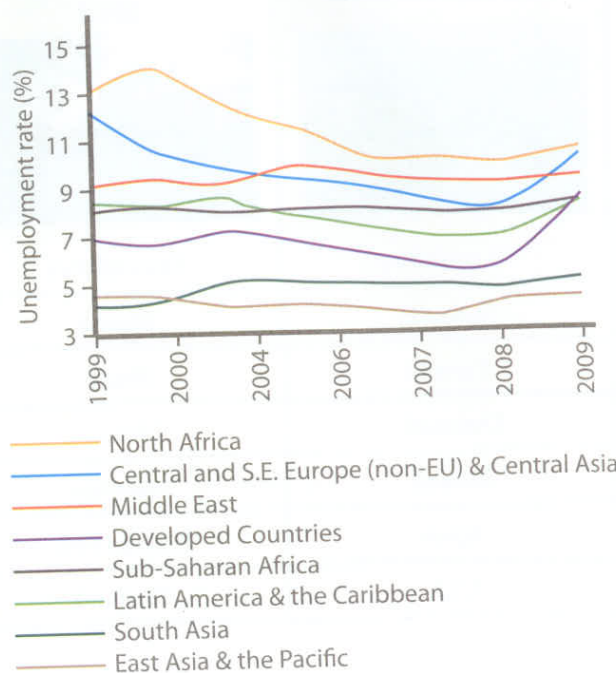
Figure 79.4 Unemployment in developing countries (1999 – 2008)

Even the United Nations Population Fund (the confusing acronym of which is UNFPA – maybe whoever came up with it also started the DNA[5]) admits that there is no 'commonly acceptable definition of underemployment' and that estimates are in fact guesstimates. The ILO reckons that 4 out of 10 people holding a job in LDCs suffer from underemployment.[6]

Agricultural sector

Linked to both low productivity, low incomes and high levels of both unemployment and underemployment, is the large proportion of the labour force working in the agricultural sector. Low productivity is strongly linked to underemployment when people in rural areas simply cannot find work and stay on marginal land which is highly unproductive. While the majority of LDCs will have a third or more of the labour force working in the agricultural sector, less than 5% is the norm in high income countries. Figure 79.5 looks at the proportion of the labour force in agriculture using the same selection of eight countries.

[5] National Dyslectic Association
[6] UNFPA, *Population dynamics in the Least Developed Countries...* 2011

Figure 79.5 Proportion of the labour force in agriculture (2010)

UNDP classification	Percentage of labour force in agriculture (2010)
Low income	
Uganda	80%
Ethiopia	80%
Lower-middle	
Indonesia	38%
Sudan	80%
Upper-middle	
Algeria	14%
Mexico	13%
High income	
Australia	3.6%
Denmark	2.6%

(Source: http://www.tradingeconomics.com/ unemployment-rates-list-by-country and World Factbook at https://www.cia.gov/library/ publications/the-world-factbook)

Informal economy

When economists speak of **formal markets**, they are referring to markets which are part of the system which acts within the boundaries of the institutions of competitive rules, tax regulations and overall legislative frameworks. All activities are within legislative guidelines and recorded in some way: minimum wages and labour taxes are paid; expenditure taxes such as VAT are registered; and social welfare contributions are paid.

The **informal sector** – sometimes referred to as the *unofficial economy* – is composed of two (often overlapping) parts in developing economies:

1. **Subsistence agriculture** and manufacturing of tools and clothing for the household, and

2. **Parallel markets** for goods which vary from cottage manufacturers of cloth and agricultural produce to drugs, weapons and prostitution. To be blunt, there is even trade in human beings – slavery – in parts of northern Africa and Asia.

Do not be fooled into underestimating informal markets; in developing countries the informal sector is a most important element in creating jobs and output. The informal sector comprises between 25% and 40% of all economic activity.[7]
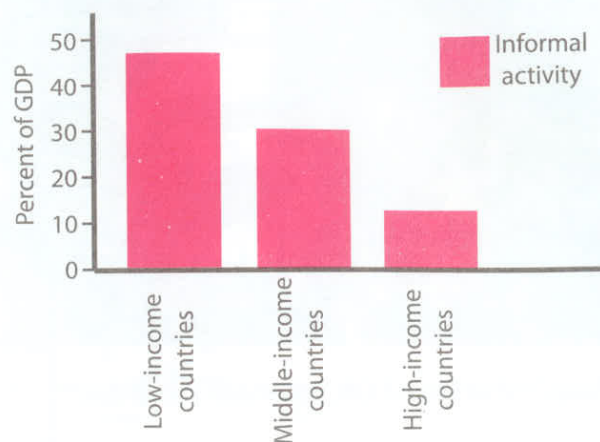
Figure 79.6 Proportion of the economy in the informal sector (2005)
(Source: WDR 2005, page 63)

The mechanisms behind informal markets are simple: market activity will arise parallel to formal markets due to complex, costly and **overly-formalised regulation** and **hazy property rights**. Add to this a weak system of administration/regulation/ supervision and an almost inevitable result is a large informal economy. Developing countries are often **dual economies**; there is a formal and informal sector side-by-side. Any number of countries will have a most astonishing difference between the two, with highly industrialised and technologically advanced urban areas just a few kilometres from rural people living off the land and trading goods with others.

The informal market can be termed a 'natural' market as it arises in answer to the different needs of people looking to satisfy different wants. The success in this is of course that each transaction facilitates the satisfaction of needs; for cassava, wooden bowls or home-made satellite dishes.[8] The informal market allows economic activity to take place that would otherwise be lost due to weak institutions and therefore in fact increases allocative efficiency since there would otherwise be very little in the way of markets.

---

7 World Bank at http://wbln0018.worldbank.org/HDNet/ HDdocs.nsf/0/3d876a6c6f40389d85256a7100589c85?OpenD ocument&ExpandSection=2

8 True! I've seen such dishes made of aluminium cans cut open and joined together. Very cool indeed.

DEVELOPMENT

"Taxes? What taxes? Oh, you mean mordidas!"[9]

However, there are a number of notable economics costs associated with informal markets:

- **Loss of income/output** in the economy. The formal economy will have difficulty in growing due to the **'unfair' competition** from an informal economy which is not paying taxes – and the informal economy will have severe restraints in moving towards and becoming part of the formalised economy due to lack of access to capital.

- An economy which has a large informal section will provide a **poor tax base** and limit potential tax revenue for the government – and set limits to merit/public goods and transfer payments.

- People working in the informal sector will have little or no **social security** or social welfare benefits.

- The poor who operate businesses within informal markets will have **little protection** from the law. Nor will they be able to insure their goods or enforce contractual agreements in a court of law. This puts both their assets and often their physical well-being in the hands of corrupt officials and gangs providing 'protection'. The payments for both amount to a business tax for which they receive no societal goods in kind.

In order to reduce the size of informal market activity the opportunity cost ratios between illegality and legality must be changed. In other words, the **benefits** in registering a business and paying fees/taxes **must outweigh the costs**. While it is in no way an easy task, economists have increasingly put forward suggestions of: better paid tax collectors with incentive-based wages to reduce corruption; reforms in legislation making it easier and cheaper to start and run businesses; and simplified tax systems shifted towards easy-to-collect tax bases.

Income inequality

"Trickle-down economics is the idea that if you feed the horse enough oats eventually some will pass through to the road for the sparrows."

JK Galbraith

While economic growth cannot be conclusively shown to influence income distribution, the opposite holds broadly true; low levels of income inequality can be shown to have beneficial effects. In addition, highly unequal distribution of income commonly has a number of generally adverse effects on human development:

- High income inequality is strongly associated with high poverty levels. A study by the International Monetary Fund (IMF) shows that **absolute poverty can be decreased** if economic growth takes place under conditions of increasing (or at least unchanging) income equality.[10]

- When there is gross discrepancy between the rich and poor, **crime levels** are measurably higher. This will have a far greater effect on the poor than on the rich.

- Inequality leads to **sub-optimal economic performance**. At high levels of income inequality, **savings** will be low and therefore so will investment, since poor people will save a significantly lower proportion of income. While the rich might save more in absolute terms, the proportion will be small. Additionally, a large proportion of expenditure in upper-income groups will be spent on imported luxuries, which does not fuel domestic production. There will also be a proportion of **capital flight**.

- High inequality enables the rich to control and exploit the political and economic system by favouritism and

---

9 A 'mordida' is, literally, a 'little bite', e.g. a bribe to the police. It is one of the most well organised systems in Mexico, where illegal businesses pay a daily "rate" of circa MXN50 (USD4) to the police to be left alone.

10 *Tax Policy for Developing Countries, International Monetary Fund* March 2001 at http://www.imf.org/external/pubs/ft/issues/issues27/

crony capitalism. Such economic activity does not focus on 'fairness' in taxes or redistribution of income via transfer payments and public/merit goods. In effect, the poor are neglected and development suffers.[11]

It appears that inequality in income distribution is something which is quite stable in many countries over time. This belies the theory of 'trickle down' economics, which advocated a 'traditional path' of free market conditions benefiting the well-to-do section of society in the belief that the incentives to produce and profit would generate income which would ultimately trickle down to the poor. I should mention that this theory has been largely discredited.

Corruption

Closely linked to weak institutions (e.g. legal system) and large informal markets is corruption, found at many levels and in many guises in administrative systems and government offices which are supposedly controlled by rules and regulations. Anyone who has been stopped by a police officer and fined for some made-up offence has seen the result of corruption first hand. Or a mother who has to pay extra for the doctor's secretary to make an appointment; a builder who needs to grease the palms of several layers of bureaucrats responsible for permits; or when important documents have been 'mislaid' only to magically surface once a gratuity (= supplemental bonus, e.g. bribe) has been paid. Less petty examples of corruption are when government officials 'skim off the top' and keep government tax funds for themselves, and often transfer the monies to banks abroad – known as *capital flight*. These and many other forms of bribery and betrayal of peoples' trust constitute corruption, which is when people acting in an official capacity of trust and responsibility misuse their position for private gain. In keeping with my propensity to call an ace of spades an ace of spades; it is theft.

Corruption is both the **cause and effect** of poor governance and corruption levels are far higher in developing countries than in developed. Lack of good governance creates fertile ground for corruption, as the mechanisms to enforce proper institutional behaviour are all too often missing. Lack of democracy, little transparency in government institutions and a media which is pro-government or simply state-controlled will all yield poor public insight into misdoings and advance corruption. This in turn will reinforce corruption in the system.

A yearly index of corruption by *Transparency International*, a non-government organisation, shows that there is clear

11 See for example HDR 2003 page 16

correlation between development and corruption. Figure 79.7 shows that development, measured by the HDI, indeed renders less corruption.
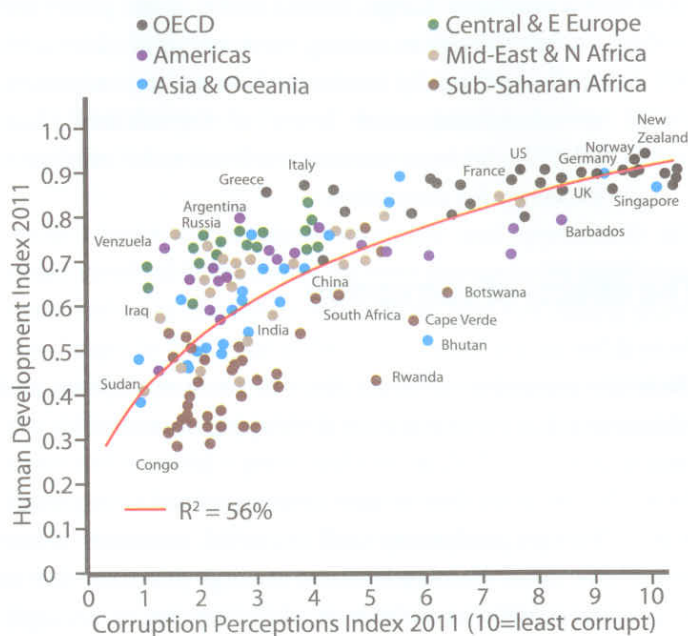
Figure 79.7 Corruption index and HDI (2011)

The causes of corruption

So why does corruption arise and what promotes it? A number of clear causal variables can be identified:

- **Complicated and extensive regulations**: When there are many licenses, permits and other legal documents necessary in, say, starting a business or opening a shop, the authorities will have a form of monopoly power to hinder entrepreneurs. If there are several papers needed and several stamps necessary to start a business, there is a clear incentive for bureaucrats to 'lose' papers or simply withhold stamps in order to get a bribe. There is also an incentive for the applicant to want to speed up the process.

- **Taxation system**: Unclear and complex tax laws mean that honest citizens will spend a great deal of time and money on filling in income tax forms correctly. Why not just 'pay' the tax inspector to fill in the forms?

- **Discretionary power**: When officials in public office have unlimited decision making power, say in granting land rights for building hotels, then there will be a clear incentive for building contractors to pay a bribe to get favourable terms for the use of private land, so-called zoning laws. There is also an incentive to bribe public

563

officials to get attractive contracts such as building municipal housing.

- **Bureaucracy**: Large bureaucracies with poor and unclear decision-making systems will inevitably lead to politically motivated hiring – e.g. nepotism, cronyism, and other forms of favouritism. Close relations between bureaucrats creates the conditions necessary for corruption.

The effects of corruption

There are a number of direct links between corruption and economic/social performance in developing countries:[12]

- Corruption **discourages investment** and thus growth. Foreign companies tend to avoid countries where the costs of doing business are high and risky due to corruption, in addition to which they are increasingly unwilling to 'play the baksheesh (bribe) game' on account of increased shareholder pressure and demands on operational openness. For example, the consultant firm PriceWaterhouesCoopers estimates that Russia loses close to $US10 billion per year in foreign investment due to corruption and a fragile legal system.[13] Paolo Maoro from the IMF has calculated that a country improving its corruption rating from 6 to 10 (10 being the least corrupt) investment can be expected to rise by an average of 4 percentage points and increase GDP per capita growth rates by 0.5%.[14]

- When firms are committed to doing business by bribery there is going to be a larger degree of 'hidden bookkeeping' and **tax evasion**.

- When countries are run by dictatorships, or an elite group, there will often be '**crony capitalism**', where attractive government contracts and soft loans go to family members or trusted friends. This has often led to highly inefficient use of government funds.

- Bribery in effect is **another tax** on doing business and the poor are hit the hardest since piecemeal bribery will have a much larger effect on those with small incomes. The real tax in the form of bribery is thus a regressive tax. An increase in the corruption index by one point (10 point scale) has been estimated to be equivalent to a 5% increase in the tax rate and decrease foreign direct investment by 8%.[15]

- Corruption has a tendency to distort government policies and skew public funds away from areas rendering general social benefits. Highly corrupt countries tend to **under-spend in human capital** goods such as education and health care. For example, a parliamentary committee in the Philippines in 2002 estimated that corruption cost close to $US2 billion annually – double the monies allocated to the education budget.[16] **Environment** issues also tend to rate poorly in corrupt societies as officials accept bribes not to enforce environmental laws. **Crime rates** are high due to ineffective and corrupt law officials.

- Corruption breeds **distrust** and scepticism towards government agencies and makes it harder to implement democratic reforms, as those who hold power will seek to uphold a system which empowers them.

- When corruption results in misuse of funds and capital flight the country can suffer from an **increase in the burden of debt**.

Poverty levels

Poverty was dealt with in Chapter 55 which showed that – unsurprisingly – the main factor in high poverty levels is low income. The aggregate estimate by the World Bank in 2011 showed some 1.44 billion people in the world living on less than USD1.25 per day – 2.6 billion if the daily income is set at less than USD2 a day. The annual World Development Report from the world bank has consistently found strong correlation between rising incomes and falling poverty rates. Figure 79.8 shows just how strong this association is where China, with an average growth rate of over 8% has decreased the proportion of the population living on less than USD1 a day by almost 50% over a 20 year period.

---

12 See the excellent paper on corruption by *Susan Rose-Ackerman, The Challenge of Poor Governance and Corruption*, at www.copenhagenconsensus.com

13 *Global Agenda Magazine* at http://www.globalagendamagazine.com/2003/petereigen.asp

14 Paulo Mauro; "Corruption: Causes, Consequences, and Agenda for Further Research", at http://www.worldbank.org/fandd/english/0398/articles/010398.htm

---

15 Ibid.

16 *Mugabe Stands Out Among the Politically Corrupt, While Banks and Energy Sector Top Dirty Business Deals Uncovered in 2002*, Transparency International press release 17 December 2002.

DEVELOPMENT
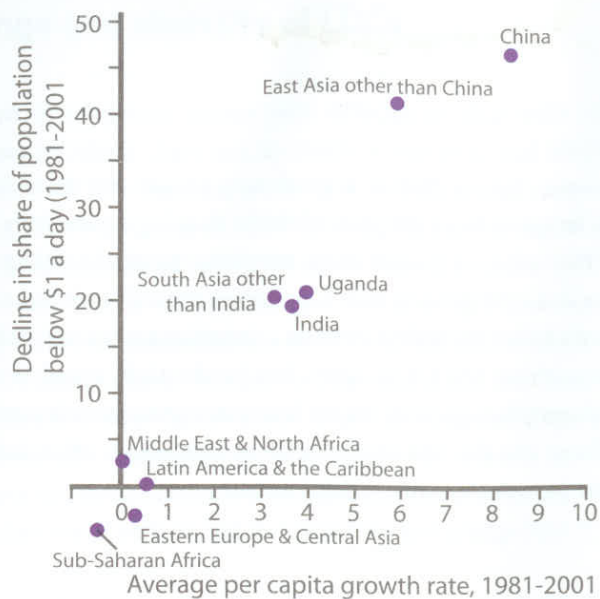
Figure 79.8 Growth and poverty reduction (1981 – 2001)   (WDR 2005, page 31)

Birth rates, death rates and population growth

Birth rates (fertility rates) and death rates (mortality rates) are usually expressed as 'live births/deaths per 1,000 inhabitants'. The birth rate minus the death rate give the population increase/decrease disregarding net immigration. Birth rates, death rates and population growth will be considerably higher in LDCs than in MDCs.

- Low income countries in 2012 had an average birth rate of 3.3% and an average death rate of 1.1; natural population growth was thus an average 2.2%.

- High income countries for the same year had a birth rate of 1.2% and a death rate of 0.84; natural population growth was 0.36%.[17]

It's not rocket science; the proportion of the world's people living in low income countries has been increasing for decades. While there is considerable debate as to the exact cause of falling birth rates in MDCs – increased incomes, improvements in social welfare systems and education have all been put forward as possible causes – there is little to debate about in terms of the raw data for populations. High population growth together with falling death rates in low income countries is a clear demographic trend.

---

17 But in fact populations in high income countries increased by some 0.8% due to net immigration. See http://www.tradingeconomics.com/high-income/birth-rate-crude-per-1-000-people-wb-data.html and World Bank online.

Urban sprawl (Todaro model)

The renowned development economist Michael Todaro has put forward a model which tries to explain why there is such heavy **migration to urban areas** in spite of the fact that all evidence shows how urban unemployment is very high. Between 2000 and 2010 the proportion of the population in low income countries living in urban areas increased from 27.55 to 31.6%.[18]

Todaro looks at the opportunity cost and opportunity gain issues involved in the choice between staying in the rural sector and moving to the urban sector. Posit that you simply expect your wages to be higher in the city – would you move? It would of course depend on how strong your expectations were and the difference in wages. It's a bit like a lottery; if one stands a chance of winning a great deal one might be prepared to pay a higher price for the lottery ticket. So too, reasons Todaro, will people function: as long as the expected wages in urban areas are considerably higher – or just assumed to be considerably higher – people will flock to urban areas. The possible gains of employment and good wages make up for the high costs of moving.

Todaro posits therefore that a *rational decision* could be made to move to an urban area as long as the expected earnings are higher than the costs of moving. The expected earnings of migration are not only wages, but the benefits of health care, opportunities and the 'bright lights/big city' atmosphere. These possible gains must be weighed against the costs of moving, such as the money costs of transport, breaking up from family and home, loss of crops and land – and the risk of simply not finding a job.

Poverty trap and pro-development cycle(-s)

In Chapter 55 the so-called 'poverty trap' was brought up: low income → low savings → low investment → low productivity → low income…etc. While avoiding the irksome "…yeah but how do you get into the game?!" issue, Figure 79.9 shows a *possible* breaking of this cycle. Let us assume a benevolent and functioning government, good financial institutions and a reasonable level of technological and human development. Increased growth yields both increased consumption, savings and taxes – the latter two leakages facilitate increased investment and government spending respectively. Increased physical and human capital increase productivity in the economy and generate growth. Now we have a positive feedback loop.

---

18   World Bank online

Figure 79.9 Pro-development cycle – and some trade-offs
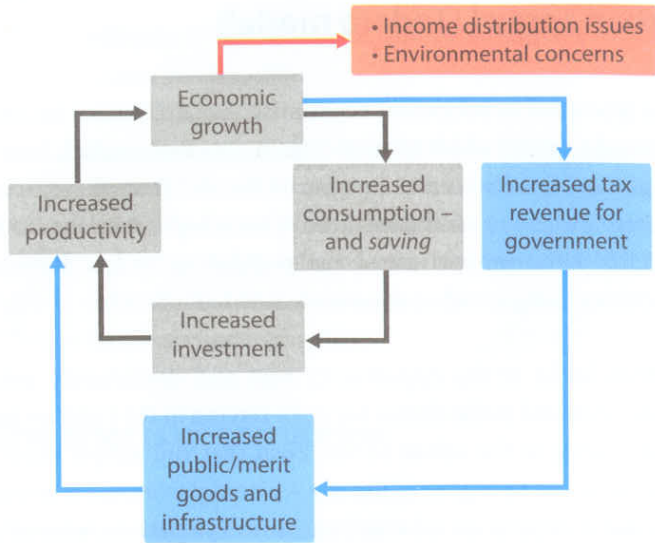
Of course, Figure 79.9 is a great way to use my mate Jeff's favourite quote from the gym when he does leg squats with 220 kgs; "Hey man, if it were easy, everyone would be doing it!" Where does the growth start – government will have scant tax revenues to set the ball in motion…incomes are low so savings and investment too are meagre…low productivity coupled with high population growth means low per capita growth rates… etc. There are also the unintended consequences of widening income gaps and environmental issues.

Figure 79.10 HDI and other data, selection of countries 2009 – 2010 (Source, HDR 2011, page 164)

HDI Rank		POPULATION							ECONOMY					
		Total (millions)		Average annual growth (%)		Urban (% of total)	Median Age (years)	Dependency ratio (%)	GDP per capita (PPP$)	Foreign direct investment net inflows (% of GDP)	Net official development assistance received (% of GDP)	Remittance inflows (% of GDP)	Public expenditure on education (% of GDP)	Total expenditure on health (% of GDP)
		2011	2030	1990-1995	2010-2015	2011	2010	2011	2009	2009	2009	2009	2006-2009	2009
117	Guyana	0.8	0.8	0.1	0.2	28.7	23.8	58.2	3,240	7.1	8.5	12.5	8.1	8.1
118	Botswana	2.0	2.3	2.7	1.1	61.8	22.9	57.2	13,384	2.1	2.5	0.7	10.3	10.3
118	Syria	20.8	27.9	2.8	1.7	56.2	21.1	67.1	4,730	2.7	0.5	2.6	2.9	2.9
120	Namibia	2.3	3.0	3.1	1.7	38.6	21.2	65.9	6,410	5.3	3.6	0.1	5.9	5.9
121	Honduras	7.8	10.7	2.6	2	52.2	21	68.3	3,842	3.5	3.3	17.6	6.0	6.0
122	Kiribati	0.1	0.1	1.5	1.5	44			2,432	1.7	15.6	6.4	12.2	12.2
123	South Africa	50.5	54.7	2.4	0.5	62.2	24.9	53	10,278	1.9	0.4	0.3	8.5	8.5
124	Indonesia	242.3	279.7	1.6	1.0	44.6	27.8	47.8	4,199	0.9	0.2	1.3	2.4	2.4
125	Vanuatu	0.2	0.4	2.8	2.4	26	20.6	70.8	4,438	5.3	16.5	1.0	4.0	4.0
	…													
143	Kenya	41.6	65.9	3.1	2.7	22.5	18.5	82.1	1,573	0.5	6.1	5.7	4.3	4.3
174	Ethiopia	84.7	118.5	3.3.	2.1	16.8	18.7	79.2	934	0.8	13.4	0.9	4.3	4.3
169	Sudan	44.6	66.9	2.6	2.4	40.8	19.7	76.7	2,210	4.9	4.6	5.5	7.3	7.3
96	Algeria	36	43.5	2.2	1.4	67.1	26.2	45.8	8,172	2.0	0.2	1.5	5.8	5.8
57	Mexico	114.8	135.4	1.8	2.1	78.1	26.6	54.1	14,258	1.7	0.0	2.5	6.5	6.5

DEVELOPMENT

Range and diversity of LDCs

I find development issues very difficult to deal with in the 'aggregate' since there are so many exceptions and statistical outliers in this field of study. For example, the reasonably straightforward issue of GDP per capita in PPP terms gives wildly different HDI values for countries with similar incomes; Botswana with a HDI ranking of 118 and GDP per capita of USD13,384 is under Guyana at 117 and USD3,240. Indonesia and Vanuatu have similar GDPs and HDI rankings but Indonesia is over 1,200 times larger in terms of population. Additionally, Indonesia consists of 18,000 islands spanning 20% of the circumference of the globe…Vanuatu has 80 islands you can reach with an inflated tyre ring and a large kite.[19]

You get the picture; lumping 'developing countries' together is like the American side of my family asking about 'European food'.[20] I end this chapter with a brief look at diversity in developing countries.

History

The six developing countries exemplified within this chapter were not chosen entirely randomly. I wished to be able to put forward that these very divergent nations in fact share a trait (or background) very common to LDCs; they were all colonised by Europeans. This has left a stark imprint in the institutions, culture, language, infrastructure and ruling elite classes left behind after emancipation during the 1960s to 1980s.

There are numerous studies linking the so-called 'colonial heritage' of Sub-Saharan Africa to the development failures during the latter part of the 1900s: European rulers subdivided the continent, artificially creating countries which erupted into civil strife and dictatorships when the colonial powers left; colonial powers created artificial elites to act as figurehead powers and this perpetuated tribal and ethnic strife within the countries; many of the poorest countries turned out to be landlocked nations with no access to sea ports; and huge amounts of raw materials and wealth were taken abroad rather than being re-invested in the host country.

Geography and climate

Jeffrey Sachs, of Millennium Goals fame, has put forward for decades how geography and climate have major impacts on development. Another famous economist, William Nordhaus, has written extensively about climate and the effects of climate change.[21] Summarising their positions in two very brief points:[22]

1. Many of the poorest countries have discernable *climatic disadvantages* which clearly work against development. Malaria and a number of other diseases limit population spread, as do harsh climates and lack of water in many instances. Drought prone areas work against agricultural stability and a lack of dependable year-round navigable water systems limit transports.

2. Huge distances and being *landlocked* result in poor transportation networks and communications infrastructure and have a resounding negative impact on trade and the costs of doing business.[23] Populations of advanced civilisations have historically often 'clustered' in coastal or river regions and this provided the groundwork for trade patterns which enabled the spread of technology and knowledge over time.[24]

Resource endowment

Resources endowment means to be 'blessed' with certain (given) resources. Fisheries, forests, natural wild life preserves, soils, fresh water, minerals and energy sources are all forms of the wealth an economy can be endowed with. Norway is one of the richest countries in the world (HDI ranked number 1 in

19 One island is called Matthew. I think I might claim it as mine. Another is called Pentecost, where they invented bungee jumping - with **in**elastic bands. A third, Tanna, is home to the John Frum Cargo Cult - the true 'Economics of Hope'.
20 Yes, it's true. I try to explain that there are slight, subtle and virtually imperceptible differences in the various cuisines of Europe – say, the difference between haggis or deep-fried Mars bars (Scotland) and börek or biber dolma (Turkey).
21 You heard it here first: I hereby predict that the 2012 or 2013 Nobel Prize in economics (more correctly the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel) will go to Nordhaus for his work on the economics of global warming.
22 See Sachs' work at www.jeffsachs.org and Nordhaus' at http://gecon.yale.edu. Also strongly recommended – and lighter! – reading is Jared Diamond's *Collapse: How Societies Choose to Fail or Succeed*.
23 There are 15 landlocked countries in Africa. All but diamond-rich Botswana can be found in the bottom 28 countries (out of 187) in the 2011 HDI rankings. See http://unstats.un.org/unsd/methods/m49/m49regin.htm
24 One of my history professors at Uppsala university in Sweden (yes, I am in fact an historian originally) explained the Viking trade routes and subsequent links to other cultures by taking a map of Scandinavia and superimposing a picture on top where all the (blue) waterways became black. "What do you see?" he asked. "Yes, you see roads! Water did not separate. Water linked!"

DEVELOPMENT

2011) by way of being the second largest exporter of natural gas and forth largest exporter of oil in the world; other large oil and gas producers such as United Arab Emirates, Brunei and Qatar (HDI rankings 30, 33 and 37 respectively); Botswana produces almost a third of the world's gem-quality diamonds in the world and has had one of the highest growth rates in the world for 30 years (HDI number 118); Australia, Chile and Malaysia are all major exporters of primary goods (HDI numbers 2, 44 and 61 respectively)…and some 20 other countries are included in a list of countries which have to a large extent managed to attain economic growth by way of natural wealth. Yes, there is quite a disparity between the HDI rankings in the resource rich countries above and they were picked intentionally to show, well,

… read on…

Natural resources both generate and sustain growth, thus an economy with an abundance of natural resources should fall in the high-income bracket, right? Well, no; quite the opposite in many cases. In fact, in so *many* cases and for such a long time that one can comfortably use the term 'The resource curse' or the 'Paradox of plenty' amongst economists and they will know what you are talking about.[25] Using an extreme example, the Democratic Republic of Congo is one of the wealthiest countries on the planet in terms of natural resources; it accounts for a third of the world's diamonds, half of all cobalt and has huge deposits of copper and tin. It has had the world's lowest growth rate and eighth lowest GDP for 40 years. At the other end of the spectrum, Singapore has basically no natural resources – importing 40% of its water from Malaysia – and has had growth rates over the past 30 years which enabled it to go from just over half of USA's GDP per capita to 120% in PPP terms.[26]

Let's have a look at the data. Figure 79.11 shows the growth rates of resource-rich and resource-poor developing countries over a 35 year period. From the 1980s onwards, growth rates in resource-poor countries have exceeded that of the resource rich. In Figure 79.12 we can see that there is in fact negative correlation between natural resource endowments (measured by exports) and growth rates.
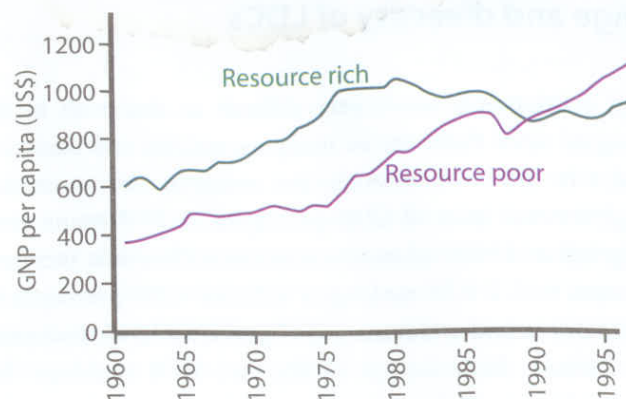
Figure 79.11 The paradox of plenty – natural resource endowment and growth

(Source: Do natural resources limit global economic development? Lancaster University, Dept of geography, 2002)
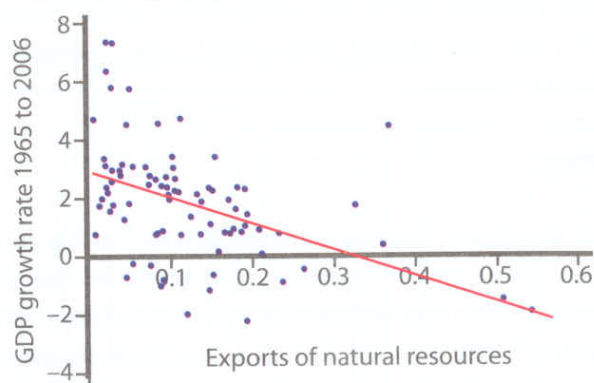
Figure 79.12 Correlation between natural resource exports and growth – 87 countries 1965 to 2006

(Source: Norwegian government report at http://www.regjeringen.no/en/dep/ud/documents/nou-er/2009/nou-2009-19-2/12.html?id=572262)

Once again a warning finger is raised; do not confuse correlation with causality! It is not the case that an abundance of natural resources has *caused* lower growth rates. Instead it is other – coincident – variables which are *linked* to both resource richness and low growth. Out of a hideously complex debate in this area, two key points might shed some light on the issues:

- Discoveries of oil/gold/gas lead to rapid increase in income and inflation. Inflation causes across-the-board rises in export prices and in fact makes all other industries less competitive. This is the so called 'Dutch disease' named after the rampant inflation in Holland during the 1960s when natural gas deposits off the coast and subsequent exports caused the Dutch Guilder to appreciate making other Dutch exports less competitive. There was in fact a net negative effect on aggregate demand.

---

25 See for example OECD publication *Natural Resources and Pro-poor Growth* 2009 and *Natural Resource Abundance and Economic Growth*, Jeffrey D. Sachs and Andrew M. Warner, Center for International Development and Harvard Institute for International Development, Harvard University, November, 1997

26 See International Monetary Fund on-line at http://www.imf.org/external/pubs/ft/weo/2006/01/data

DEVELOPMENT

- *Weak institutions* – primarily democratic and legal – provide an environment where warring factions, ethnic power struggles and gangsterism flourish when valuable resources are discovered and opportunities arise for the unscrupulous/corrupt/strong.[27]

Political systems and political stability

It has long been demonstrated that economic institutions such as legal systems guaranteeing property rights, regulatory institutions upholding competition and financial institutions such as banks all have a major influence on economic growth and development. Many of the struggling developing nations have all been de facto dictatorships until fairly recently – Colombia, Malaysia, Indonesia, Mexico and Zaire to name but a few. A goodly number are still dictatorships; China, North Korea, Burma, and Sudan all lack universal voting rights and are either one-party states or run by a military group. Of 195 countries in the world, 48 are considered 'not free' and an additional 60 'partially free'.[28] Virtually all of the 48 pegged as 'not free' have had reoccurring periods of political turmoil (often together with civil war) during the past 20 years.

Political instability means that society is laced with **insecurity** and **fear**. What is the incentive to plant and build fences on land you don't own, when there is a very real risk that marauding gangs – supported by the government's blind eye – come by to 'tax' you? What if civil unrest keeps you ever-ready to flee to the next province or country? Or that your chances of prosperity, and even survival, depend on being a member of the right tribe or clan – the one in power at the moment?
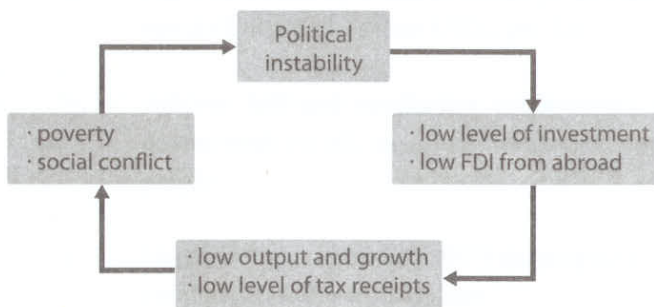
Figure 79.13 Political instability and development

The results (Figure 78.4) are painfully obvious; economic activity will be very low as will both domestic and foreign investment. The effect is perhaps more serious in case of the latter, as foreign firms would be able to provide much needed injections of capital to developing countries – not to mention the transfer of knowledge and technology. There is understandable reluctance to invest in a country which cannot guarantee the safety of personnel, capital or profits.[29] The absence of this vital growth mechanism leads to a vicious cycle, this time of political instability → low investment → poverty → political instability.

Political instability goes a long way to explain the plight of Africa during the past 30 years; one third of all countries in the continent were involved in armed conflicts – most of them civil conflicts – and all of these nations will be found in the bottom third of the Human Development Index.[30] Investment in Africa, both domestic and foreign, is the lowest in the world, and an estimated 40% of the continent's wealth is held abroad. It is the only continent to grow poorer between the late 1980s and early 2000s. One bright light is Botswana which has had stable governance during the same period, giving it one of the highest per capita growth rates in the world during the past 30 years.[31]

There are three broad aspects of **political stability** which help foster the development process:

1. **Implementation of policies and long term goals:** Stable governments will be more prone to plan for the future, formulate long term goals, and implement them. It is also easier for countries with political stability to attract foreign investment and aid.

2. **Rule of law:** Stability helps in setting down and upholding laws which govern property rights, competition laws, and starting businesses – all of which help economic growth. There must be a functioning **police**, **judicial**, and **penal system** where police act on authority limited by law; cases are tried and rulings are based upon law so as to exclude arbitrary rulings; and those who are convicted of crimes have the right to appeal their case. The rule of law prohibits the oppression of the weak by the strong. A good system of law is one where the rules

---

27 Look up 'blood diamonds' or 'conflict diamonds'. Steel yourself for some of the most gruesome chapters in human history.

28 Annual rankings done be Freedom House, see http://www.freedomhouse.org/sites/default/files/inline_images/FIW%202012%20Booklet--Final.pdf

29 The issue of crime, corruption and lack of clarity in the legal system is something frequently discussed by many of the expatriate parents in my school in Mexico. During the international financial/liquidity crisis of 2008/09 I spoke to several parents who were going to be relocated to other – safer – countries as the crisis exacerbated (= intensified) violent crime.

30 *Coping with conflict*, The Economist, Jan 15th 2004

31 *First get the basics right*, The Economist, Jan 15th 2004

DEVELOPMENT

apply equally and to all – rulers and ruled alike. The rule of law will provide stability and reliability since citizens will be able to rely on fair treatment in the system. This is particularly important in matters concerning the protection of property rights and competition, since both are necessary for economic growth.

3. **System of power transfer:** Whatever the level of autocracy/theocracy/democracy, a system of governance which has shown stability over time will probably have a lower degree of social disruption such as civil war and armed conflicts.

The last aspect warrants a final comment. *Political instability and conflict* go hand-in-glove; internal wars and conflicts, ethnic divisions and violence are both **causes** and **effects** of instability. The correlation between conflict and development is very clear: high levels or frequency of conflict are strongly correlated with low human development. Of all countries experiencing conflict since 1990 we find:[32]

- Nine out of ten of the lowest HDI countries

- Seven out of ten of the lowest GDP per capita countries

- Nine out of ten countries with the highest infant mortality rates

- Eight of the ten countries with the lowest primary school enrolment ratios

Summary & revision

1. **Common characteristics** of LDCs include:

   a. *Low productivity* and per capita *incomes*

   b. *High unemployment* and *underemployment* levels

   c. Large *agricultural sector*

   d. Large *informal economy* as a proportion of GDP

   e. *Unequal income* distribution

   f. Relatively high levels of *corruption*

   g. High birth and death rates – far greater *population growth* than MDCs

   h. Migration to urban areas (*urbanisation*)

2. Developing countries show a broad **spectrum of diversity:**

   a. *History*

   b. *Geography* and *climate*

   c. Differences in *resource endowment*

   d. Political systems and *stability*

32 HDR 2005, page 154

DEVELOPMENT

80. The Millennium Development Goals

Key concepts:

- The Millennium Development Goals (MDGs)
- Evaluation of the MDGs

The Millennium Development Goals (MDGs) are specified targets which the international community agreed upon in 2000. The core issue was to halve poverty and improve the welfare of the poorest nations by the year 2015. All 191 members of the United Nations pledged to meet these eight goals by 2015. The goals were intentionally phrased as clearly measurable targets and are monitored by the World Bank and the International Monetary Fund (IMF). The attainment levels of each goal are exemplarily updated at http://mdgs.un.org/unsd/mdg.

The Eight Millennium Development Goals (MDG)

I've never met a teacher who can remember more than five so I shall be brief here. Each goal below is phrased as a 'target'. You may see that the focus on 'measurability'.

1. **Eradicate extreme poverty and hunger**. Targets:

   a. Halve (between 1990 and 2015) the proportion of people living on less than USD1 per day

   b. Achieve full and productive employment and decent work for all, including women and young people

   c. Halve (between 1990 and 2015) the proportion of people who suffer from hunger

2. **Achieve universal primary education**. Targets:

   a. Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling

3. **Promote gender equality and empower women**. Targets:

   a. Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015

4. **Reduce child mortality**. Targets:

   a. Reduce by two thirds, between 1990 and 2015, the under-five mortality rate

5. **Improve maternal health**. Targets:

   a. Reduce by three quarters the maternal mortality ratio

   b. Achieve universal access to reproductive health

6. **Combat HIV/AIDS, malaria and other diseases**. Targets:

   a. Have halted by 2015 and begun to reverse the spread of HIV/AIDS

   b. Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it

   c. Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases

7. **Ensure environmental sustainability**. Targets:

   a. Integrate the principles of sustainable development into country policies and programmes and reverse the loss of environmental resources

   b. Reduce biodiversity loss, achieving, by 2010, a significant reduction in the rate of loss

DEVELOPMENT

c. Halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation

d. By 2020, to have achieved a significant improvement in the lives of at least 100 million slum dwellers

8. **Develop a global partnership for development**. Targets:

a. Develop further an open, rule-based, predictable, non-discriminatory trading and financial system

b. Address the special needs of least developed countries

c. Address the special needs of landlocked developing countries and small island developing States

d. Deal comprehensively with the debt problems of developing countries

e. In cooperation with pharmaceutical companies, provide access to affordable essential drugs in developing countries

f. In cooperation with the private sector, make available benefits of new technologies, especially information and communications
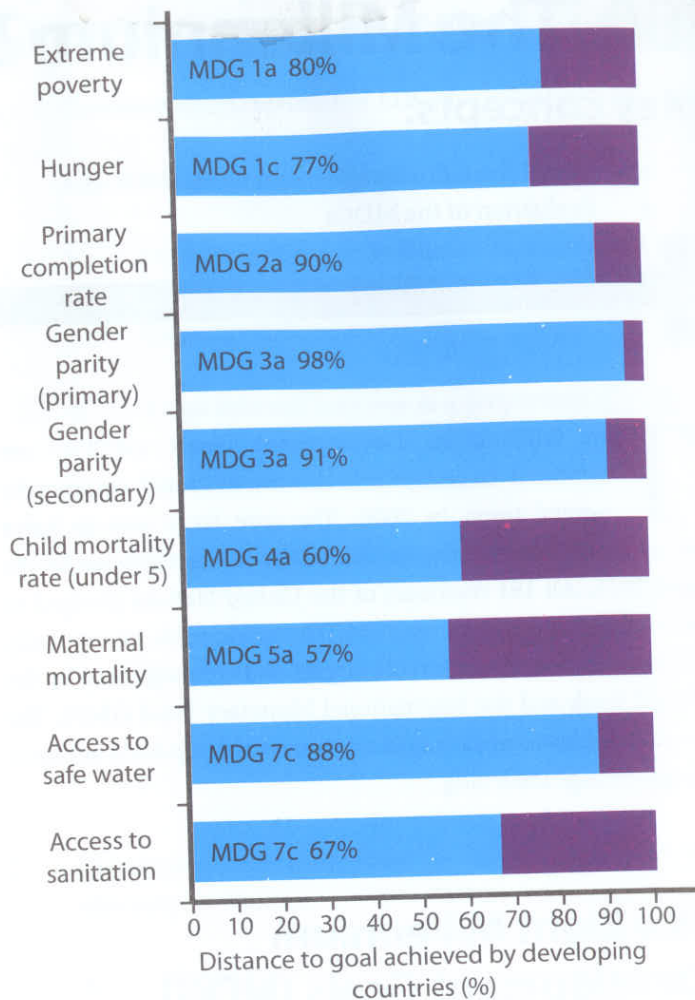
Figure 80.1 Attainment levels of MDG goals 2011

Evaluation of the MDGs

According to the World Bank, roughly two thirds of LDCs were on track as of 2011 to meeting the targets set down in the MDGs. Major successes include the probable achievement of gender parity in primary and secondary education and access to clean drinking water. It also appears that the reduction in poverty will be close to targets. The least successful attainment levels are in health, primarily child and maternal mortality rates. Figure 80.1 shows the 'distance' to goal attainment as a percentage of total attainment.

Summary & revision

The eight **Millennium Development Goals** are

1. Eradicate extreme *poverty* and *hunger*.

2. Achieve universal primary *education*.

3. Promote gender *equality* and empower women.

4. Reduce child *mortality*.

5. Improve *maternal health*.

6. Combat HIV/AIDS, malaria and other *diseases*.

7. Ensure environmental *sustainability*.

4.2

81. Measuring Development – Single Indicators

Santiago, Chile - a rapidly developing economy

Key concepts:

- Problems of measuring development
- GDP and GNI per capita
- Purchasing power parity (PPP)
- GDP and GNI adjusted for PPP
- Health indicators
- Education indicators

Problems of measuring development

As can be seen throughout this section, development is far from a universally agreed-upon measurement and can have a number of broad meanings in different contexts. Perhaps the only point of agreement amongst economists is that no single measure will suffice to either define or measure development. Instead, one often outlines a set of criteria in order to arrive at some agreed-upon common ground – giving a semblance of objectivism in the use of measurements of development. However, it should be noted that simply by using the term 'development' we are in fact leaning heavily on a value-judgement, i.e. that development is something to strive toward ('good') while lack of development is to be avoided ('bad'). In reality things are not always that simple as many people in 'newly developed' areas would tell you. As my friend Anna Collins-Falk points out – after spending over 20 years working for development projects in Africa and South East Asia – it is not always that someone (e.g. a development organisation) actually *asks* the local people what they want…

In any case, the task of measuring development will come up against many obstacles. Here is a compilation of the main problems in measuring development, most of which can be put within two headings: **definitional** problems and **accounting/ measurability** problems.

1. Definitional problems

a. **Growth** is not development, as shown in some depth earlier. Simply focusing on economic development disregards any number of social indicators which are increasingly regarded as essential in measuring the overall well-being of citizens.

DEVELOPMENT

573

b. **Social indicators** such as longevity, literacy and level of democracy attempt to adjust for this problem, yet there is as yet no agreed-upon all-encompassing definition of development. The definitional problems in outlining development give us perhaps the greatest measurement difficulties, since it becomes so hard to pin down quantitative variables in what is evidently a qualitative area of study.

2. Accounting and measurability problems

Poor statistics of lagging and unreliable quality, black markets, parallel markets, unaccounted-for income, (barter and subsistence level agriculture), etc. mean that averages do not show distribution of income and exchange rates skew the purchasing power of incomes.

a. Poor **statistical reliability** and quality. Many countries do not have the wherewithal to compile accurate statistics on growth and development, and the time involved in collecting data and processing it will cause lags in data series. This is immediately evident when looking through official statistics for a number of countries and the problem is frequently noted by institutions such as the UNDP and the World Bank.

b. Accounting for income as an element in overall development gives a number of **weaknesses of GDP/ GNP figures** as observed earlier in this section; black markets, barter economies and non-market activity such as subsistence farming are not shown and will skew (= slant) any national income figures.

c. In accounting for and using indicators of development there are a number of **negative side effects** which are difficult to factor-in to the wider development picture. Negative externalities such as pollution are difficult to measure as are environmental impacts such as deforestation.

d. Many **other indicators** of societal well-being are conspicuously absent. For example, personal safety issues such as robbery, assault and other crime are not included in any of the most commonly used development indices.[1] None of the measurements used above show,

for example, how many people are allowed to vote freely in elections; availability of culture and free time; level of corruption necessary in dealing with government officials; freedom of education, movement and choice for women; access to information and free media; lack of crime...

GDP and GNI per capita

Recall GDP and GNI from Chapter 37; GDP is the money value of output *within* a country's borders and disregards foreign ownership while GNI (GNP) adds on *net property income* from abroad. Linking also to the balance of payments in Chapter 70; foreign direct investment (FDI) outflow from the domestic economy is accounted for in the financial account and repatriated profits from foreign ownership flow back in on current account. Conversely, FDI inflows to the domestic economy create outflows of repatriated profits. The net flow over a period of time of repatriated profits is the *net property income from abroad*. Ponder for a minute right here – where would the majority of FDI flows be heading to/from in terms of LDCs and MDCS – and thus will LDCs predominantly have a *net positive* or *net negative* flow of repatriated profits? Then ask yourself if GNI should be greater than GDP in LDCs.

The answer can be seen in Figure 81.1 where the previously used UNDP groups of countries show that LDCs have lower GNI than GDP – fully expected when taking into consideration that in most cases one can expect larger FDI flows to LDCs (from MDCs) than the opposite. (In fact, if you flip back to Chapter 79 and Figure 79.1, every country in the table that is not classified as a high income country has lower GNI than GDP.) Somewhat surprisingly, the lowest income countries show a *slightly* higher average GNI per capita but the general pattern is quite clear; MDCs are net investors abroad while developing countries are net recipients of FDI.

---

1   At the risk of getting targeted for 'sanction with extreme prejudice' by the Mexican Tourism Board, I must tell the following enlightening story. Like all private schools in Latin America, I had several students in my class from rather wealthy families. One day a very good student stayed behind class at my school in Mexico and since the IA deadline was that lesson I simply waited for the "my computer ate my homework" excuse. "Em, Matt, I need an extension..." I replied, truthfully, that I simply don't do extensions. Tears welled up in the student's eyes; "I know...but my sister has just been kidnapped..." The deadline was extended. (The sister is safe!)

UNDP classification	GDP per capita (2010)	GNI per capita (2010)	GNI minus GDP
Low income			
Afghanistan	501	410	–91
Cambodia	795	750	–45
Average for the group (35)	526	529	3
Lower-middle			
Indonesia	2,946	2,500	–446
Sudan	1,425	1,270	–155
Average for the group (56)	1,716	1,622	–94
Upper-middle			
Algeria	4,567	4,390	–179
Mexico	9,133	8,930	–203
Average for the group (54)	7,264	5,886	–1,378
High income			
USA	47,153	47,340	187
Denmark	56,245	59400	3155
Average for the group (70)	36,781	38,745	1,964
Average for world	46,287	46,782	4951

Purchasing power parity (PPP)

We were sitting outside having a cold beer at a nice restaurant in Oshakati, Namibia in November 2002: Anna, Per, myself and John Shamhula – our fearless guide/nanny. John was curious about Sweden and asked many questions about our lives there and I think we all were amazed not at how different we were but how alike we were! At one point John asked me; "How much would this beer cost in your country?" He held up his half-litre stein for emphasis. After some quick head counting based on an exchange rate of one Swedish Crown (SEK) to 0.9 Namibian dollars (NAD) and a few murmurs of agreement from the colleagues I said "About NAD 40 at a place like this. Now you know why we think that NAD 5 for a large beer here is so cheap!"

John got furious – and loud: "Forty dollars for a beer?! For ONE beer?! This... this... this is CRIMINAL! These people are thieves! They should... they should be *punished*! Punished most

severely!" What John disregarded, brilliant mathematician that he was, was that translating Swedish Crowns into Namibian dollars meant that a great deal was lost in translation; *while SEK 1 was equal to NAD 0.9 according to the official exchange rate* (or NAD 1 = SEK 1.1) you could buy a great deal more for 100 SEK – i.e. NAD 90 – in Namibia than in Sweden. The exchange rate made my Namibian colleagues look very poor in comparison as the differences in price levels did not show in the **purchasing power** of each currency.[2]

A simple purchasing power example

The trick is to convert the two currencies in some way to show the relative power for each citizen in his/her own country. What we did, sitting there in the restaurant trying to cool John down, was to figure out how many beers you could buy for an average teaching salary (Figure 81.2):

Figure 81.2 Income (teaching salary) and PPP, Sweden vs. Namibia (2002)

Country	Average monthly salary for a teacher	Price of a ½ litre beer	Beers per salary
Sweden	SEK 18,000	SEK45	400
Namibia	NAD 2,000	NAD5	400

In other words, the exchange rate grossly distorted (underestimated) the actual purchasing power of Namibian teachers. Since the average Swedish teacher in 2002 earned nine times as much as the Namibian teacher, one might be led to assume that the Swede could buy nine times as much. This was clearly not the case, at least for beer, since the half-litre stein was

2 During my job interview with the headmaster of the school I worked at in Mexico, there was a brief silence in the conversation whereupon the headmaster asked "Well, do *you* have any questions?" "Yes indeedy," I replied, "What does a beer cost in a restaurant in Mexico?" There was a loooooooooong pause. I could imagine the head whispering to the head of secondary; "...this guy's nuts...possibly alcoholic...". To the head's credit, I was given an answer and quickly figured out that the offered salary increased my purchasing power (well, for beer at least) by a factor of three. I use restaurant prices for beer as my PPP indicator since the good is domestically produced, contains an element of service costs and is heavily taxed. It turns out to be a pretty good indicator of relative purchasing power – far superior to the Big Max Index at least. My soon-to-be boss happily conceded the point and hired me...and since my colleague in the econ dept didn't touch alcohol I figure between the two of use we had on average one *normal* economist – which might be a contradiction in terms.

nine times as much in Sweden. In other words, the purchasing power of Namibian and Swedish teachers in terms of beer was identical!

At an equitable 'beer exchange rate', the Namibian and the Swede would have to be able to buy the same quantity of beer in the other's country; the 'beer exchange rate' at purchasing power parity would be when SEK18,000 = NAD 2,000, i.e. SEK 1 = NAD 0.055 or **NAD 1 = SEK 18**. Since we were getting sixteen times this amount (by getting NAD 0.9 for each SEK) when we changed money at the banks, the exchange rate was stacked highly in our favour in terms of purchasing power.

What we need to do in order to make international comparisons of income more realistic – e.g. compare apples with apples – is to adjust the average incomes to show what one can actually buy for the money in the country in question. Basically we are asking "How much is USD100 worth in New York or New Delhi?" Ask yourself also if GDP/GNI figures for LDCs are consistently under- or overvalued in terms of what simple exchange rate data shows.

GDP and GNI adjusted for PPP

The UNDP does a purchasing power calculation for all countries every year by putting all the world's GDP figures into a common currency, the US dollar, and adjusting the figures to show what these dollars would buy a citizen in a given country. These PPP adjusted figures are more frequently used instead of exchange rate based figures. Figure 81.3 shows the GDP per capita in US dollars at both current exchange rates and adjusted for PPP for selected countries in 2006.

Figure 81.3 GDP and GNI per capita adjusted for PPP – selected countries 2010 (current USD) (Source: World Databank on line)

UNDP classification	GDP per capita in PPP (2010)	Percentage increase (+) or decrease (-) compared to non-adjusted values	GNI per capita in PPP (2010)	Percentage increase (+) or decrease (-) compared to non-adjusted values
Low income (unadjusted values from Figure 81.1 in brackets)				
Afghanistan	(501) 1,207	+140%	(410) 1,060	+159%
Cambodia	(795) 2,193	+176%	(750) 2,080	+177%
Average for the group	(526) 1,278	+143%	(529) 1,305	+147%
Lower-middle (unadjusted values from Figure 81.1 in brackets)				
Indonesia	(2,946) 4,325	+47%	(2,500) 4,200	+68%
Sudan	(1,425) 2,255	+58%	(1,270) 2,030	+60%
Average for the group	(1,716) 3,573	+108%	(1,622) 3,673	+126%
Upper-middle (unadjusted values from Figure 81.1 in brackets)				
Algeria	(4,567) 8,432	+84%	(4,390) 8,100	84%
Mexico	(9,133) 14,563	+60%	(8,930) 14,400	+61%
Average for the group	(7,264) 10,080	+39%	(5,886) 9,972	+69%
High income (unadjusted values from Figure 81.1 in brackets)				
USA	(47,153) 47,153	+/-0	(47,340) 47,340	+/-0
Denmark	(56,245) 40,162	-29%	(59400) 41,100	-31%
Average for the group	(36,781) 37,292	+1.4%	(38,745) 37,318	-3.7%

In adjusting GDP figures to assess the purchasing power of citizens in each country, some remarkable patterns emerge. One of the notoriously most expensive countries in the world, Denmark, has significantly lower purchasing power than national income figures would indicate – PPP values are almost a third lower than simple exchange rate values indicate. Simply put, rich countries' GDP figures are often *overestimated* using conventional exchange rate methods, while developing countries show the opposite. It is often the case that the price level in LDCs is much lower than in developed countries – which means that the domestic LDC currency will go a lot further at home than in an (expensive) MDC. Using PPP methodology to compare GDP will render higher per capita income for LDCs than only using an exchange rate to convert the figures.

Hence, Denmark's GDP per capita is 29% lower when adjusted for PPP while Cambodia's turns out to be 177% higher. Low income countries see their PPP income increase by over 140% compared to unadjusted values and this trend is noticeable at lower-middle and upper-middle income levels. Only at high income levels do PPP values fall and this is, of course, to be expected. Another comparison shows that, as a percentage of USA's GDP per capita, average income in Denmark is not 19.3% higher but rather 13% *lower* when PPP is used instead of the foreign exchange rate. The data conforms nicely to economic theory in that *lower income countries' GDP and GNI values are consistently underestimated by simple exchange rate figures* – and that the lower the income of a country the greater the underestimation.

Health and education indicators

There are numerous indicators for health and education but I have saved the 'standard' ones (life expectancy at birth and mean/expected years of schooling) for the iteration on the Human Development Index (HDI) in Chapter 82. Figure 81.4 shows the divergence between high and low income countries – but also within the four classifications, for example Cambodia and Afghanistan are both classified as low income countries but Cambodia spends more than 6 times as much as Afghanistan on education measured as a percentage of public expenditure.

Figure 81.4 Health and education indicators – selected countries 2010 (or latest year in parentheses)

UNDP classification	Physicians per 1,000 people	Public health expenditure (% of total govt expenditure)	Primary completion rate (%)	Expenditure on education (% of total govt expenditure)
Low income				
Afghanistan	0.2 (2005)	1.6	35 (2005)	NA
Cambodia	0.2	10.5	87	NA
Average for the group (35)	0.2	NA	57.9 (2005)	NA
Lower-middle				
Indonesia	NA	7.7	96 (2005)	26%
Sudan	NA	9.7	40 (2002)	NA
Average for the group (56)	0.78	5.4	83 (2005)	NA
Upper-middle				
Algeria	1.13 (2002)	8.1	96	NA
Mexico	1.5 (2002)	12.1	99.5 (2005)	NA
Average for the group (54)	1.7	9.5 (2002)	97.6 (2005)	15.6 (2002)
High income				
USA	2.3 (2002)	22.3	100	13.7 (2005)
Denmark	2.9 (2002)	16.8	100	15.7 (2005)
Average for the group (70)	2.8	16 (2005)	97 (2005)	12.6 (2005)

DEVELOPMENT

Honestly, the figures in Figure 81.4 speak for themselves but right about now you might start thinking along the lines of an economist; what is the correlation between these and other development indicators?! No, I do not intend to do the Anglo tick-box-fill-in-the-blanks for you. Get yourself pen and paper and start figuring it out.

Other indicators of development (some interesting suggestions here!)

Once again, the main problem in outlining welfare in developmental terms is solving the problem of measuring what in effect cannot be objectively measured. In order to overcome this problem, economists have tried to construct measurements based on relatively quantifiable and identifiable variables which can be used to show relative living standards and overall human welfare. The discussion on measuring development has so far been limited to national income, which is too limited in scope to be used as the only indicator. Here are a few other common development indicators:

- the 'distance' between the richest 10% and poorest 10% in society

- amount of medical doctors per 1,000 inhabitants – and average travel time to a hospital

- the number in a population have access to clean water and sanitation

- average daily caloric intake per person

- road miles and other transportation networks in km per capita

- telephones/computers/internet access per 1,000 inhabitants

- infant mortality rates measured by infant deaths per 1,000 newborns

- percentage of under-14s active in the labour force

- GDP/GNP adjusted for economic 'bads' such as pollutant levels and environmental damage and the addition of non-money activities such as barter

- adjusting income figures to show long run sustainability by factoring in the depreciation of **environmental capital**, e.g. natural resources

Over the years I have enjoyed putting the following question to my students: "Come up with your own – measurable! – development indicators which can be correlated with 'mainstream' indicators used in textbooks." I give you a brief list below together with an encouragement to do the numbers yourself, i.e. find the figures for one of the indicators below for a selection off countries and correlate them with, say, GDP per capita or enrolment in tertiary education.

- Amount of people in prison as a percentage of the population

- Ratio of government spending on education to spending on military goods

- Ease of obtaining handgun permits or amount of handguns per household on average

- Corruption level in national postal services (one of my all-time favourites)

- Amount of police per capita (some rather interesting results here)

The advantage of any of the above is that they are all *measurable* in purely quantifiable terms. The disadvantage is that no single indicator will suffice to show overall development. The solution is often to create a composite (= multiple, combined) index where a number of different variables are used together and indexed in order to make comparisons possible. The most commonly used index for development levels is the *Human Development Index* looked at in the next chapter.

The WaterAid Cracker Challenge

ilikemyquality1 Subscribe 1 video

4,000 Children die every d...

4000 children die everyday because...you keep writing on their fucking food.
SERFL242 3 months ago

391,363

DEVELOPMENT

Postscript

Yes, I too see that the math doesn't quite add up! Technically, world GDP should equal world GNI (since net property income flows for the world will equal zero) but we seem to be getting FDI from Mars. The answer is simply that there are some very large discrepancies in the income figures available for many of the LDCs in the list. Total world per capita GNI minus GDP is off by USD495 – which in fact is a discrepancy of just one percent.

Summary & revision

1. **Measuring development** come up against two problems:

   a. *Definitional* problems

   b. Accounting and *measurability* problems

2. **GDP** will in many cases be **larger than GNI** in LDCs since the latter takes into account repatriation of profits. GNI is GDP plus net property income from abroad and many LDCs will see greater net outflows than net inflows.

3. **Purchasing power parity** (PPP) means that income values are adjusted to show the actual spending power of a US dollar (as most international comparisons are done in US dollars).

4. When GDP and GNI are **adjusted for PPP**, low income countries will mostly have higher PPP incomes than at simple exchange rate values. High income countries will mostly see PPP values be lower than exchange rate values.

5. **Health indicators** include infant mortality rates, life expectancies, doctors per 1,000 inhabitants, availability of potable water and health expenditure as a percentage of GDP or total government spending.

6. **Education indicators** include percentage of children in school, number that complete primary/secondary/tertiary education, percentage of population that is literate and spending on education as a percentage of GDP or total government spending.

82. Measuring Development – Composite Indicators

Key concepts:

- Composite indicator – HDI
- HDI figures for LDCs and MDCs
- Other composite indicators
- Comparing GDP/GNI per capita and HDI values

"Scandinavia comes out on top according to the HDI because the HDI is basically a measure of how Scandinavian your country is." Bryan Caplan, American right-wing economist[1]

Simply put; there is no 'best single indicator' for standard of living or development. Economists have for years compiled aggregated data and compiled *composite* indicators – i.e. 'mixed together' several key indicators to give a wider reach of measurement.

Composite indicator – HDI

A common failing in our strange science is the implicit notion that for something to be truly definable it must be measurable. While this is not quite true it is often helpful to have a set of objective standards which can be applied to different phenomena in order to compare them. The Human Development Index is an attempt to create a way to rank countries according to their degree of development using three different – measurable – variables and then composing the results in a single value. It is published annually in the Human Development Report from the United Nations Development Programme (UNDP).

The HDI uses four indicators to show health, education and standard of living:

1.   **Life expectancy** at birth; measures health

2.   a.   **Mean years of schooling** for 25 year olds, and

        b.   **Expected years of schooling** for children entering school; measures education levels

3.   **GNI per capita** adjusted for purchasing power; measures standard of living

The scores are put together in a composite index where 1.00 is the highest attainable value. This is the value used in Figure 82.1.

This young Fijian Islander can expect at least 8 years of education, a health service and (maybe) a democratic government.

---

1 I have to admit that even though I am Scandinavian (Swedish) I simply cannot take offence to this statement. It's too true.

HDI rank	Human Development Index (HDI) Value	Life expectancy at birth (years)	Mean years of schooling (years)	Expected years of schooling (years)	Gross national income (GNI) per capita (constant 2005 PPP $)	GNI per capita rank minus HDI rank
Human Development Index groups						
Very high human development	0.889	80.0	11.3	15.9	33,352	0.918
High human development	0.741	73.1	8.5	13.6	11,579	0.769
Medium human development	0.630	69.7	6.3	11.2	5,276	0.658
Low human development	0.456	58.7	4.2	8.3	1,585	0.478
Regions						
Arab States	0.641	70.5	5.9	10.2	8,554	0.643
East Asia and the Pacific	0.671	72.4	7.2	11.7	6,466	0.709
Europe and Central Asia	0.751	71.3	9.7	13.4	12,004	0.785
Latin America and the Caribbean	0.731	74.4	7.8	13.6	10,119	0.767
South Asia	0.548	65.9	4.6	9.8	3,435	0.569
Sub-Saharan Africa	0.463	54.4	4.5	9.2	1,966	0.467
Least developed countries	0.439	59.1	3.7	8.3	1,327	0.467
Small island developing states	0.640	69.6	7.3	10.8	5,200	0.675
World	0.682	69.8	7.4	11.3	10,082	0.683

Figure 82.1 Overview of HDI values for the world (Source: HDR 2011, table 2)

HDI figures for LDCs and MDCs

As shown in Figure 82.1, the UNDP divides countries into groups; low, medium, high, very high and Scandinavian human development. No, kidding of course – checking to see if you read the footnotes. In any case, Figure 82.2 shows a selection of countries and their respective rankings according to the HDI 2011. I have used the previous eight countries and added the highest and lowest HDI ranked countries.

HDI classification	HDI value	HDI rank	GNI per capita 2011 (PPP)	Income classification (see Ch 81)
Very high human development				
Norway	0.943	1	47,557	High
USA	0.919	4	43,017	High
Denmark	0.895	16	34,347	High
High human development				
Mexico	0.77	57	13,245	Upper-middle
Medium human development				
Algeria	0.698	96	7,658	Upper-middle
Indonesia	0.617	124	3,716	Lower-middle
Cambodia	0.523	139	1,848	Low
Low human development				
Sudan	0.408	169	1,894	Lower-middle
Afghanistan	0.398	172	1,416	Low
D.R. of Congo	0.286	187	280	Low

Figure 82.2 HDI and GNI per capita (PPP) for 187 countries (Source; HDR 2011, table 1)

Comparing GNI per capita and HDI values

Ploughing through the Human Development Report yields pretty much the results seen in Figure 82.2, namely that there are considerable differences between pure income rankings and development as measured by the HDI. In the small sample of eight I have used consistently in this section, we see that:

- The top HDI ranked countries retain their ranking from the income classification done earlier.

- Two countries move down in the HDI rankings compared to income rankings; Algeria and Sudan both move down a notch.

- One country, Cambodia, moves up a notch.

This is consistent with what the HDR has been pointing out for years now, namely that many of the resource rich countries – read 'oil', frequently – often have high GDP/GNI per capita figures but that this does not always translate into equitable distribution of wealth/income, services such as health care and education across a broad sector of society, nor government policies for taxing and redistributing the income. Two countries in our tiny sample have moved down in the rankings. It turns out that Algeria, the world's 15th largest gas and oil producer, sees 45% of GDP and two thirds of government revenues coming from gas and oil and that 98% of all export revenues from natural gas and oil.[2] Sudan is in the middle of turmoil that might well lead to all out civil war over rich oil deposits in South Sudan, which broke away in July 2011.

Looking at a few extreme discrepancies between GNI per capita and HDI rankings, we indeed see that – with a few exceptions – the 'oil curse' rears another head here:[3]

- *The* major exception, obviously, is Norway, the 13th largest oil producer in the world and HDI rank number 1. Other exceptions would seem to be United Arab Emirates, Brunei, Qatar, and Bahrain, all large oil/gas producers which come in at HDI values of above 0.8 and are categorised as high human development. It bears pointing out, however, that Qatar (HDI rank 37) has a GNI per capita that is four times higher than Australia's (HDI rank 2).

- Equatorial Guinea (HDI rank 136) has a GNI per capita of USD17,608 – higher than Estonia at 34th place in the HDI rankings. There are huge oil and gas deposits in the country.

- Kuwait is the 10th largest oil producer in the world with a GNI per capita of USD47,900 (almost identical to HDI number 1, Norway) and comes in at an HDI ranking of 63.

Other composite indicators

A number of other indicators have been put together in order to measure how well off a country is and also how well off a country might be in the future. Here is a short list of some composite indicators being used by development organisations:

- **Human Sustainable Development Index** (HSDI): Adds in a sustainability dimension to the HDI in order to see future living standards and consumption possibilities. The HSDI adds in a fourth metric, namely CO_2 emissions.[4]

- **Human poverty index** (HPI): This index uses three basic measurements of deprivation. The first relates to survival and measures by using the probability of a citizen not surviving to the age of 40. The second deals with knowledge and measures the percentage of the adult population that is illiterate. The third measures standard of living.

- **Multidimensional Poverty Index** (MPI): Replaced the HPI in 2010. It is similar to the HDI in that it measures health, education and standard of living but instead of four metrics it relies on ten – including child mortality, malnutrition and access to electricity and potable water.

- **Inequality-adjusted Human Development Index** (IHDI): The IHDI was introduced in the Human Development Report in 2010 to show discrepancies in distribution of the components of the HDI. Basically if there is perfect distribution (zero inequality) then the HDI will be the same as the IHDI. The higher the level of inequality, the more any given HDI value will be downgraded in the IHDI. For example, the US at HDI

---

2 International Monetary Fund at http://www.imf.org/external/pubs/ft/survey/so/2011/int012611a.htm

3 See HDR 2011, tables 1 and 2. Oil and gas figures taken from Indexmundi at www.indexmundi.com

4 See http://ourworld.unu.edu/en/the-2010-human-sustainable-development-index/

ranking number 4 winds up at 23 while Sweden at HDI ranking 10 winds up at number 5.[5]

Summary & revision

1. The **Human Development Index** (HDI) is a composite indicator of the level of development a country as attained. It measures health, education and standard of living by using four indicators:

    a. Life expectancy at birth

    b. Means of schooling years and expected years of schooling

    c. GNI per capita in PPP

2. HDI values are between zero (lowest) and one (highest. The **UNDP classifies** countries into four levels:

    a. **Very high** human development – between 0.793 and 0.943, such as Norway, Australia and Singapore

    b. **High** human development – between 0.698 and 0.783, such as Cuba, Bulgaria and Brazil

    c. **Medium** human development – between 0.522 and 0.691, such as Thailand, Egypt and Botswana

    d. **Low** human development – between 0.286 and 0.510, such as Zimbabwe, Nigeria and Haiti

3. **Comparing GDP/GNI** per capita figures **with HDI values** often clarifies that all income per capita figures are averages and that weak governments, poorly functioning institutions and bad policies often negate high income levels.

4. Other composite development indicators include:

    a. The Human Sustainable Development Index

    b. The Human Poverty Index

    c. The Multidimensional Poverty Index

    d. The ~~Scand...~~ Inequality-adjusted Development Index

---

5   Suspiciously enough, all the Scandinavian countries (except number one Norway) go up in their rankings by between 4 and 7 places. It seems that the HDI is indeed a measure of how Scandinavian a country is!

4.3

83. Domestic Factors and Economic Development

Health Care in a remote area. The Royal Flying Doctor service at Coober Pedy in South Australia

Key concepts:

- Education
- Health care
- Use of appropriate technology
- Access to credit and micro credit
- Empowerment of women
- Income distribution

Education

An investment in knowledge pays the best interest. Benjamin Franklin

One of the highest rates of return in both economic growth and development issues is undoubtedly to be found in education. Every country in what we now call the developed world experienced the brunt of growth and development once primary education had been established. In fact, one of the world's most famous development economists, Nobel laureate Amartya Sen, points out that Japan as early as 1870 had higher literacy rates than Europe – allowing rapid transition to an industrialised power 50 years later. Sen points out that the 'miracle' economies of East Asia had similar levels of literacy.[1]

The benefits of education are both **private** and **social**. The element of merit good in education is very high which, to a

large extent, ends the discussion on 'private or public provision' in developing countries. The societal return in education investment is so high that a majority of developing countries strive to spend as large a portion of government funds as possible on the education sector. The OECD countries spend an average 6.2% of GDP on education while some of the poorest countries in the world spent an average of 6.4%.[2] One of the few success stories to be had in Sub-Saharan Africa during the past 30 years, Botswana, spent an annual average of 10.3% of GDP on education between 2006 and 2009.3

Private and social benefits of education

- An individual stands to benefit in increased lifetime *earnings*, higher *productivity* and better *jobs*. Studies show that each additional year of education increases individual output by 4 – 7%, while increasing lifetime earnings by 10 – 20%.[4] The benefits to society are **growth** and an *expanded tax base*.

- In addition to the economic benefits, being able to read and write will greatly expand the **quality of life** for a person who is able to take an active part in society

1 *Development as freedom, Amartya Sen*, page 41

2 See HDR 2011, page 165

3 HDR 2011, table 10

4 *Teach a child – transform a nation, Basic Education Coalition* report using UNDP, World Bank and UESCO statistical material, pages 7 – 8

via media. This is put most directly by Amartya Sen, development economist and Nobel laureate, in that the personal benefits of education "…exceed its role as human capital in commodity production." Education enables people to read, communicate, form serious opinions and ultimately argue their case. This furthers political participation and studies show strong links between education and political stability/democracy.

- Individuals' **health** will improve due to knowledge in sanitation and how to prevent transmissible diseases, amongst them HIV/AIDS. The particularly gruesome figures on the spread of HIV amongst young people lends particular weight here; the World Bank estimates that 15 – 24 year olds account for 60% of all new HIV infections in developing countries.[5] Education improves health care simply by enabling people to read food and medicine labels, pamphlets on health care and family planning and so forth. The Swedish author Henning Mankell who spends most of his time in Africa puts it rather bluntly; "All the statistics available are clear – it's the ones who can read that have the lowest prevalence of AIDS".[6]

- There are few areas in development studies where the correlation between two variables is as strong as that which is shown between increased education of women and falling birth rates.[7] Education greatly **empowers women** in giving them an opportunity to provide for themselves. They will have safer and healthier pregnancies, and will provide better nutrition and living conditions for their children. Increased financial independence is perhaps the best overall form of population control, and evidence shows that women who are allowed a secondary education have significantly lower birth rates than women who are uneducated.

- Education renders **economic benefits** such as growth and increased tax revenues in the long run – allowing increased social spending on public and merit goods so vital to society. **Social benefits** are improved health, lower death rates and birth rates, and political/democratic stability.

Story-Time!
Brightest Africa

My long-standing habit of going off for long runs wherever I am in the world caused a good deal of amusement and not a little worry among my colleagues during our time in Namibia. The first time out running I lost track of time and the sun went down. Man, when the sun goes down in the Namibian bush it gets dark; no warning, no long period of dusk, just pitch black nothing. Naturally I got myself severely lost – I mean, how do you know where you came from when you can only see a few feet ahead. After getting instructions from some highly amused local people, I got home three hours late and was chastised by worried colleagues and forbidden to run later than 16:00.

On one of these late afternoon runs way out in the desert/bush, I was hailed by a "Helloooooooooo!" which I barely heard through AC/DC in the earphones. In the distance was a waving person who, upon running closer, I discovered was a very pretty young girl in a school uniform flashing a brilliant smile. She was on her way home from school, and in one hand she had a bag with books and groceries and in the other hand her shoes. She was barefoot.

We had a nice chat the way one can with neat young people in Namibia. She was most polite and very open and laughing. This is why I was able to ask a somewhat rude question; "Why aren't you wearing your shoes?"

Her answer is the reason I am wasting space here. She said "Oh no, these are for school only! It is most important to be proper in school. When I finish school I will be able to get a job and then I will wear my shoes when I take the bus to work."

I wish all my IB students had walked a few miles in her shoes.

Health care

A common characteristic of many LDCs is that health care is lacking, poor or simply non-existent. Few LDCs spend the 11%

---

5   *Teach a child – transform a nation*, page 15
6   *Jag dör men minnet lever vidare; Henning Mankell*, page 59
7   *Todaro*, pages 279 – 285

of GDP on public health care that high-income countries do; medium HDI countries average 4.5% and low HDI countries 5.1%.[8] The outcomes in terms of life expectancy and infant mortality rates are clearly seen in Figure 83.1.

Figure 83.1 Health expenditure and health indicators for selected countries 2009/'10

HDI classification	% of GDP spent on health care (2009)	Life expectancy at birth (2009)	Infant mortality (per 1,000 live births) (2009)	USD per capita spent on health care (current USD, 2010)
Very high human development	11.2	80.0	6	
Norway	9.7	81.1	3	8,091
USA	16.2	78.5	8	8,360
Denmark	11.2	78.8	4	6,400
High human development	6.7	73.1	19	
Mexico	6.5	77	17	604
Colombia	6.4	73.7	19	472
Medium human development	4.5	69.7	44	
Algeria	5.8	73.1	32	178
Indonesia	2.4	69.4	39	77
Low human development	5.1	58.7	117	
Afghanistan	7.4	48.7	199	38
D.R. of Congo	9.5	48.4	199	16

The difference between developing and developed countries in health spending is staggering. When adjusting for low income and high populations, Norway, USA and Demark spend between USD6,400 and USD8,000 per capita on health care. This is more than ten times the average of high human development nations and more than 160 times that spent by low human development countries.

The correlation between health expenditure and life expectancy is markedly high up to about USD1,500 per capita as shown in Figure 83.3. Avoiding the tricky issue of causality (again), most economists would agree that there is a strong reinforcing element in life expectancy and health care spending – e.g. that higher life expectancies result in increased health care spending and vice versa.
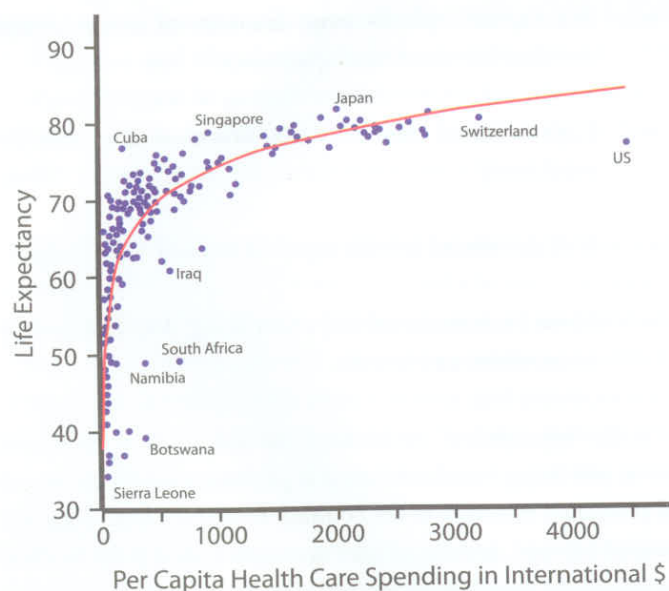
Figure 83.2 Correlation between health spending and life expectancy (Source: UNDP data 2004/2005)

Education and health care have also been shown to be strongly interlinked and **self-reinforcing**. Estimates by the World Bank indicate that a 10% rise in primary school education is associated with an average 10.8 month increase in life expectancy.[9] Better educated people will earn more money and lead longer, healthier lives. This will increase national income and improve tax bases – which in turn creates more funds for social spending on education and health care. The problem is another Catch 22, since income and tax bases are needed to provide the health care which…can create income and tax bases. (See 'Pro-development cycle' in Figure 79.9, Chapter 79.)

Use of appropriate technology

Low levels of education, infrastructure coupled with growing populations, natural resource stress and high underemployment doesn't sound like a match made in heaven for high-tech FDI does it? Development economists use the term 'appropriate' here to signify that investment and other capital-building needs to build on and take into account the internal resource endowment of the developing country. Technology might be deemed inappropriate in the following developmental contexts:

- Foreign expertise needs to install and upkeep complex machinery

- Spare parts are not available domestically but need to be imported

---

8 HDR 2011, table 10

9 *Teach a child – transform a nation*, page 8

DEVELOPMENT

- The capital replaces large amounts of cheap labour creating increased unemployment

- High levels of human capital are needed to run the machinery

- Well developed infrastructure is needed

- Meets basic sustainability criteria, e.g. does not destroy water tables and forests

As such, 'appropriate' technology would involve increased output and living standards yet still pay attention to the social and economic situation in the country. Examples[10] include solar powered laptops, mobile phones and light bulbs(!) for tropical countries with poor electricity grids and an absolutely ingenious refrigerator using evaporative cooling rather than electricity.[11]

Access to credit and micro credit

In acting as a conduit (= channel) for investment, a modern banking system provides a number of vital services in a money-based economy:

- Banks provide a system for **payment services** which is central to a market system involving supply chains of wholesalers and retailers.

- Banks enable entrepreneurs to get access to **start-up capital**. Few businesses are started using household savings alone. Banks allow entrepreneurs to realize their business ideas by borrowing against future profits. The cost of the loans – interest – is spread out over a period of time, enabling the entrepreneur to use capital now and pay for it later.

- An efficient banking system allows **profitable business** ventures to blossom, thereby allocating venture capital efficiently.

Development banks

As efficient domestic banking systems have been noticeably lacking in many LDCs, funding for small and medium

---

10 Complete – very cool! – list at http://listverse. com/2010/06/12/10-cases-of-appropriate-technology/

11 I found out about this as a proud Rolex owner when I received a news bulletin some ten years ago that this invention won the Rolex Award for Enterprise.

domestic enterprises has also been unavailable. Much of the funding came from foreign banks and went to large enterprises which could put up sufficient collateral (= security for loans). In answer to this gap in liquidity, a number of development banks have arisen in developing countries to supply long term loans. Much of the money has in turn been supplied by loans and aid monies from international development organisations and domestic governments.

Group lending schemes – micro credits

Yet the development banks still leave a considerable gap in credit availability to small scale farmers and local traders, as development banks often will not bother with loans of a few hundred – or even a few thousand – US dollars. It is simply far more time-consuming and costly to administrate 1,000 loans of USD200 than two loans of USD100,000. This credit gap is increasingly being filled by small-scale credit organisations known as *micro* credit or *group* lending schemes.

Muhammad Yunus – founder of Grameen Bank and 2006 Nobel Laureate. (Photo Ed Schipul Creative Commons License)

Like all really good ideas, it is astoundingly simple. A group of 5 borrowers band together and borrow collectively a sum of money from a development bank or other financial institution. Each member of the group is given part of the total loan and has an obligation to the group to repay his or her loan – but the entire group is responsible for the bank ultimately being paid. There are two major benefits of the system: 1) the group can borrow enough money to make it worth a bank's while; 2) internal pressure within the lending group induces members to pay their debts and interest on time. One of the most notable successes of a group lending scheme is the **Grameen Rural Bank** in Bangladesh, started in 1976. Individuals can borrow as little as USD100 within the group. This has enabled thousands

of poor people to gain access to desperately needed capital to fund small but life-supporting business ventures, such as buying cloth for producing clothing. It also turns out that over 90% of the borrowers are women and that default rates (= non-payments of loans) are less than 1%.[12]

Empowerment of women

One of the Millennium Development Goals is the promotion of *gender equality* and the empowerment of women – but it is made quite clear that this must underline all other goals since viewing any other goal, such as poverty reduction or providing universal education "without promoting gender equality will both raise the costs and decrease the likelihood of achieving the other goals."[13]

While 'gender' in fact implies both male and female, it is commonly women who are subjected to the many forms of prejudice and bias, and I limit the discussion to gender inequality pertaining to women. Gender issues play a pivotal role in human development for two reasons:

1. Development is defined as a broad, across-the-board increase in living standards and independence. Clearly, when half the population is excluded in several respects, such as legal rights and access to education, there is limited development scope. Therefore, gender equality is a measure of development in its own right.

2. Gender equality is a **facilitator of development**. For example, virtually all studies show that women's education reduces child mortality, increases growth and improves health.[14]

Education: About two thirds of the children who do not go to school are girls, for the simple reason that poor households have calculated a poor rate of return on educating girls. This is erroneous, as there is ample evidence that investment in the education of girls is a most cost-effective way of development – perhaps the most cost-effective way. Education enables women to take more control of their lives by having the opportunity to get jobs. Better educated women marry later, which lowers birth rates and increases the survival rate of children. Female education is basically an investment that raises national income

– and national income will in turn lead to more gender equality in education and elsewhere.[15] In fact, an estimated 0.4 – 0.9% of the difference in growth rates between East Asia and Sub Saharan Africa, South Asia plus the Middle East is due to gender inequality in education in the latter three areas.[16]

Legal rights: There are large differences between men's and women's right to own and inherit property in many developing countries. Having access to the family plot of land does not always mean that women will also have ownership or control of the land, as discriminatory laws, customs and practices are all too common. In not having basic property rights, women are in a most vulnerable position – they have low 'bargaining power' within the household. Women become dependent on men and will have little say in how resources are used. So while women in many developing countries shoulder most of the household chores and the tending of land for home produce, they are often unable to get credit for small investments or use the proceeds from selling excess produce.

Health care: Gender inequality in health care is abundantly clear in the statistics on child mortality for male and female children in a goodly portion of the developing world (China and South Asia); more male than female children survive. Social and cultural norms favour sons, which has led to selective abortion, neglect of female children and outright infanticide (= killing of infants). Amartya Sen, the famous Indian development economist and Nobel Laureate, famously asks "Where are the missing 100 million women in India and China?!"[17] In addition to access to health care, it is clear that gender inequality in schooling and employment increases the spread of HIV/AIDS, as women are so often powerless within the home and unable to deny the men sex.[18]

The sum of the above is that gender inequality dis-enables women from partaking in education, civil rights and health care. Furthermore it limits their ability to participate in economic, social and political structures. This in turn will have detrimental effects on the health care and education of children – which in turn will lower future national income and negate development.

---

12 *From tiny acorns, Economist* December 10th 1998; and http://www.grameen-info.org/

13 *Millennium Development Goals – A look through a gender lens,* UNDP 2003

14 See excellent overview *Does Gender Inequality Reduce Growth and Development?*; World Bank Development Research Group November 1999, page 1

15 *Gender Inequality, Income, and Growth; Are Good Times Good for Women?*; World Bank Development Research Group May 1999, page 21

16 *Does Gender Inequality Reduce Growth and Development?*; World Bank Development Research Group November 1999, summary

17 *Development as freedom, Amartya Sen,* page 104 – 107

18 *Engendering development* (summary); World Bank Policy Research Report 2001, page 9

Income distribution

The on-going debate about income distribution which seems to arrived at a 'half-way' conclusion: there is broad agreement that global income distribution has remained relatively unchanged for some 20 years at a Gini coefficient of 0.65. This translates into the top 5% richest individuals in the world getting about $^1/_3$ of total PPP adjusted income. The bottom 5% would receive about 0.2% of global income.[19] Within countries is, however, an entirely different matter. For example the US has seen how the richest 10% of households have increased their share of total income from 39% to 49% between 1967 and 2010.[20] Another extreme example of increased income inequality would be China – yes, 'would' be if the Chinese government released figures on income. Since 2001 China has refused to release figures on income distribution but every estimate done for the past 20 years indicates a Gini level that has increased from 0.41 in 2000 to around 0.47 in 2010.[21]

Income distribution and growth

While there is no conclusive evidence over longer time periods that growth *causes* increased income inequality, the evidence of the inverse is stronger; high levels of inequality can be shown to have a negative effect on both growth and the reduction of poverty.[22] The Human Development Report of 2005 puts it rather bluntly: "Extreme inequality is not just bad for poverty reduction—it is also bad for growth. Long-run efficiency and greater equity can be complementary."[23] There are three factors contributing to this correlation:

1. Relative income equality allows for a greater **sharing of societal resources** and indicates the existence of some form of redistribution system. Increased income will therefore be 'spread' via taxes to societal goods such as health care and education – which in turn is pro-growth. The 'Asian Tigers' are good examples.

2. Given a larger proportion of poor people in the economy, fewer will have the ability and collateral to be **qualified as borrowers**. Poor people simply cannot borrow against future earnings. This stifles economic activity.

3. High income inequality often results in small but **powerful elites** uninterested in 'fairness' in tax rates and redistributive mechanisms. Social spending as a percentage of GDP is therefore often lower.

4. Gross inequality – both in terms of income, gender-based policies or regions – leads to a larger section of society lacking **access to education**. As has been pointed out earlier, education is a very strong development driver.[24]

5. Inequality will disadvantage rural poor and lead to weak participation in various forms of **democratic institutions** – which undermines the development of democracy and also to internal conflicts over, for example, natural resources.

Summary & revision

1. **Education** has huge social and private benefits for LDCS; lifetime earnings increase, tax revenues grow as incomes grow, quality of life expands, it empowers women and lowers birth rates and improves health.

2. **Health care** is strongly linked to increased life expectancy, lower infant mortality and increased incomes.

3. **Appropriate technology** is technology which inexpensively increases productivity while using local production factors and resource endowment.

4. **Credit systems** and **micro-credit** enables businesses to start up and the basic form of a monetary economic system to take place.

5. The **empowerment of women** has numerous developmental benefits; higher survival rates of infants, increased incomes for women lower birth rates, legal rights allow women to take on credit and start firm, and higher status and equality will lower the spread of HIV/AIDS.

6. There is no conclusive evidence that increased income *cause* income inequality. However, the inverse seems to hold – *increased income equality is pro-growth.*

---

19 See World Bank policy paper 3865; *Global income inequality: What it is and why it matters*, 2006, page 20

20 Business Week, Oct 10th 2010, *The rich get richer…and you know the rest.*

21 See http://www.chinadaily.com.cn/china/2010-05/12/content_9837073.htm

22 See *HDR* 1996 and World Bank at http://www.worldbank.org/poverty/data/trends/inequal.htm

23 HDR 2005, page 53

---

24 HDR 2005, pages 51 – 59

4.4

84. International Trade and Economic Development – Problems in LDCs

Key concepts:

- Trade and LDCs
- Over-specialisation on a narrow range of products
- Price volatility
- Lack of access to international markets

HL extensions:

- How long-term changes in the terms of trade create development barriers in LDCs

Contrary to what the politicians and religious leaders would like us to believe, the world won't be made safer by creating barriers between people. Michael Palin of Monty Python

The UN has estimated that trade barriers which limit access of developing countries to markets in the developed world mean an annual loss of export revenue more than twice that of all aid monies received in 2000.[1] The world has increasingly moved towards trade liberalisation in many areas except those which most concern developing countries; basic commodities, agricultural goods and textiles. It has been estimated by the World Bank that the elimination of tariffs would primarily benefit developing countries – gains of some USD500 billion by 2015 would lift some 300 million people above the poverty line. It bears mentioning that most of the gain would come from developing countries removing their own barriers to trade.[2]

Trade and LDCs

The data is overwhelming in this area; developing countries which have increased trade have also seen economic growth at far higher rates than countries with low levels of international trade.[3] Yet let us be very clear and add the following: international trade is not substantially pro-developmental without good institutions and good policies in tandem. LDCs which have succeeded in attaining high growth rates by way of international trade have also had political stability, high levels of government spending on education, infrastructure and health care, and functioning policies for FDI. Many of the poorest countries in the world face considerable obstacles to free and fair trade and this has limited their development potential for decades.

Over-specialisation on a narrow range of products

It is evident that dependency on a few commodities for export revenues has been damaging for many of the poorest countries in the world, and the link between dependency on a small basket of export goods and absolute poverty is very strong.[4] In 2010 12% of the world's population living on less than $US1.25 a day in the poorest countries in the world (*least* developed countries) were dependent on a few export commodities.[5]

1   *Todaro*, page 492
2   World Bank, Global Economic Prospects (2002), page 168

3   See UNCTAD's Trade and Development Report 2006, pages 8 – 10
4   *Least developed countries report*, UNCTAD 2002, page 101
5   http://www.unohrlls.org/en/ldc/25/

DEVELOPMENT

Figure 84.1 shows five developing countries which all have more than 40% of export earnings originating from *one single agricultural commodity* and over 50% coming from three commodities.[6]

6 Notice that **cash crops** (i.e. crops which are grown for the single purpose of exportation) such as tea, tobacco and coffee make up virtually all of the commodity exports in these five countries.

Figure 84.1 Dependency on commodities – selected least developed countries Source: FAO at http://www.fao.org/DOCREP/005/Y3733E/y3733e0d.htm

Country/ territory	Export earnings of top single agricultural export commodity				Export earnings of top three agricultural export commodities		
	Percentage share in		Earnings as a percentage of GDP (1998)	Commodity	Percentage share in		Commodities
	Total merchandise exports	Total agricultural exports			Total merchandise exports	Total agricultural exports	
Burundi	75	83	7.2	Coffee (green)	89	99	Coffee (green); tea; sugar refined
Guinea-Bissau	48	91	6.3	Cashew nuts	51	98	Cashew nuts; cotton; palm oil
Sao Tome and Principe	69	97	16.9	Cocoa beans	70	100	Cocoa beans; coffee (green); copra
Ethiopia	62	69	5.4	Coffee (green)	75	84	Coffee; dry-salted sheepskin; crude org. mat.
Malawi	59	74	23.8	Tobacco leaves	70	87	Tobacco leaves; tea, sugar

It borders on an understatement so say that the least developed countries rely heavily on exporting commodities – these countries saw 80% of export revenues coming from commodities during 2007 – 2009 and there are indications that they are becoming even more dependent on this narrow range of exports.[7] The outcome for a country such Malawi is dependency and **vulnerability**[8]:

- While 'a narrow range' does not imply that all developing countries are exporting the same type of commodities, a number of the same goods are in fact produced in many countries since they have a comparative advantage in these goods – which makes it difficult for low income countries with inadequate capital markets **to diversify** into other industries.

- The difficulties in moving up in the scale of value-added production are made worse by the fact that most developed countries which import the developing countries primary goods (around three quarters of all LDC exports of primary goods) will have **differentiated tariffs** on processed goods. (See *Lack of access to international markets* further on.)

- Small scale farmers supplying global markets with their produce are often price takers without benefit of **market knowledge**. Intermediate buyers can set prices low in order to increase profit margins and this pricing power severely limits farmers' ability to get a fair market price. The most vulnerable producers in such uneven negotiations are small farmers which make up the majority of output, for example coffee, which is

7 UNCTAD report *Enabling the Graduation of LDCs: Harnessing the Role of Commodities and Improving Agricultural Productivity*, 2012, page 15

8 Look up the Singer-Prebisch thesis from the 1950s. It explains how the decline in commodity prices (even then!) relative to manufactured goods led to declining terms of trade and locked LDCs into an commodity-export trap.

grown in some 70 countries where family farms make up about 70% of output. Oxfam reckons that Ugandan coffee farmers get about 2.5% of the retail price in the UK while one of the world's largest producers, Nestlé, has a 26 to 30% profit margin on instant coffee.[9]

- Agricultural production is directed towards the export sector rather than the domestic sector. This has frequently been **linked to foreign debt** and non-convertible currencies; in order to service foreign loans the country focuses on producing cash crops for hard currency in order to pay off interest and amortise loans. Fertile lands are used for debt repayment rather than food production.

- Low PED and PES for primary goods mean that changes in demand/supply will inevitably result in price volatility, e.g. large changes in price. Such **export earnings instability** makes it difficult for producers to plan future output since revenue streams are so unforeseeable. This hampers investment in primary production and lowers growth rates.

Price volatility

The link between dependence on exporting primary products and low human development is well established.[10] The short run volatility of commodity prices makes it difficult for producers to plan output and foresee income – thus making both tax revenues and government investment in infrastructure difficult to foresee. Figure 84.3 shows just how severe these fluctuations can be.

Figure 84.2 Average yearly price fluctuations (5) for selected commodities (1986 – 1999)

Commodity	Index	Commodity	Index
Coffee	25.5%	Jute	18.1%
Sugar	22.4%	Cocoa	17.7%
Poultry	21.4%	Groundnut oil	16.2%
Rubber	18.7%	Butter	16.1%
Palm oil	18.7	Cotton	15.9

Source: Food and Agricultural Organisation (FAO) at http://www.fao.org

Over the long run average **commodity prices have fallen continuously** in real terms for over 150 years (see Chapters 9 and 77 for diagrammatic illustrations) and this has several implications for primary goods exporters.

- As exports in developing countries – primarily the least developed countries – make up for a large proportion of GDP (Figure 84.1) such nations will be most *vulnerable to world market prices* for commodities. For example, in Zambia 2008, world market prices were forced up due to the speculation of 'food scarcity' which caused this staple food to increase in price rapidly – which of course affected millions of people dependant on maize meal.[11]

- *Falling global prices* for commodities will make exporting countries worse off in terms of ability to purchase much-needed imports (HL: see terms of trade further on) and also make debt servicing harder.

- *External shocks* to the production system can be disastrous – floods, drought, earthquakes and other natural disasters will have far greater consequences for primary goods producers than for secondary and tertiary.

- *Weak infrastructure* such as transportation, storage and cooling systems can create severe difficulties for primary goods producers to get their goods to international ports and markets.

Lack of access to international markets

Developing nations accounted for approximately the same value of trade in goods as the EU in 2009 (4.7 trillion and 4.55 respectively) but all nations were hit hard by the 2008 recession. LDCs and the least developed countries saw exports of goods fall by 22% and 27%.[12] Tragically, as mentioned in Chapter 65, the rich countries responded true to form with increased protectionism. It became evident that the much-needed finalisation of the Doha Development Round of WTO talks (Chapter 63) was put on a back burner when MDCs put trade barriers to use in order to stimulate domestic demand during recession. This was nothing entirely new for LDCs, having faced severe limitations in access to developed markets for decades.

---

9 *Europe and the coffee crisis*, Oxfam 2003 at http://www.oxfam. org/eng/pdfs/pp030226_EUcoffee10.pdf

10 See for example *The least developed countries report 2002*, UNCTAD, page 137

11 UNCTAD report *Enabling the Graduation of LDCs: Harnessing the Role of Commodities and Improving Agricultural Productivity*, 2012, page 46

12 World Trade Report 2010, page 28

Tariff escalation

While the Uruguay Round of trade talks established that developed countries would successively lower tariffs towards developing countries, there have been limited effects as yet. The World Bank estimates that the poorest countries have gained the least since their exports are precisely the goods where protectionism is the highest; agriculture and textiles. This means that while average tariffs in developed countries might have fallen considerably, the *effective* (trade weighted) *average* tariff on countries exporting agricultural goods and textiles is still far higher than for manufactures. Poor people – broadly defined as those living under $US2 a day – dependent on these goods thus face more than double the tariff wall of developed countries.[13]

Developing countries often face higher tariffs on goods which have higher proportions of *value-added*, i.e. a tariff on cocoa will have lower tariffs than on cocoa butter or chocolate. For example, raw ('green') coffee entering the EU has no tariff while coffee that has been processed in some way, e.g. roasted and/or ground, faces 7.5% to 9.0% tariffs.[14] This method of **tariff escalation** reduces the demand for processed goods and inhibits developing countries' ability to move from low to high value-added. Figure 84.3 illustrates how average tariff rates for agricultural products and textiles in a developed country, Canada, are higher when imported goods have a higher element of value-added.

The tariff escalation mechanism slows down the process of diversifying into processing industries, limits the accumulation of skills and capital, slows export growth since basic commodity prices are continuously falling, and perpetuates primary goods exporters' vulnerability to global price fluctuations.

MDC subsidies

An estimated 60% to 90% of all agricultural subsidies in the US, EU and Japan go to the 25% largest firms in the agricultural sector. These subsidies have the effect (explained in Chapter 65) lowering international prices and destroying markets for LDC exporter which cannot compete with subsidised goods. I have long favoured cotton and sugar in my use of examples – using the US and EU respectively – as these two goods exemplify the hideous wastefulness and resource misallocation at a global level.

Cotton in the US: The US subsidises its circa 25,000 cotton growers to the tune of USD2 billion per year – or an estimated USD31 billion between 1995 and 2010 – 3.5% going to the largest 10 producers.[15] The WTO upheld a complaint by Brazil in 2005 that the US was illegally distorting the market but so far the US has ignored the ruling and subsidies continue. Oxfam, a large UK based non-government organisation, estimates that removal of all US and EU subsidies on cotton would raise world market prices by 10% and allocate production and jobs to far more efficient producers in North West Africa.[16]

Beets working.

Figure 84.3 Tariff escalation in Canada (Source: Market Access for Developing Countries' Exports, IMF and the World Bank 2001, page 24)

Sugar in the EU: About 40% of the entire EU budget goes to farm subsidies and one fourth of this goes to large agri-businesses – multinational companies in fact. The top earners were the sugar companies, totalling €606 million for five sugar conglomerates in 2010.[17] Apart from the huge costs to EU taxpayers, one

---

13 *Global Economic Prospects and the developing countries*, World Bank 2002, page 57

14 *Government Actions to Support Coffee Producers – An Investigation of Possible Measures from the European Union Side, Germán Calfat Institute for Development Policy and Management*, University of Antwerp, Belgium 2002, page 19.

15 http://farm.ewg.org/progdetail. php?fips=00000&progcode=cotton

16 Time, April 9 2010, *Why the U.S. Is Also Giving Brazilians Farm Subsidies*

17 See for example The Guardian, 4 May 2010, *EU sugar and dairy companies largest recipients of farm subsidies.*

must also take into account the de-developmental aspects of thousands of lost jobs in North African countries which are perfectly endowed with resources to produce sugar. In addition to this there are huge efficiency losses in subsidising sugar in the EU since sugar beets are used rather than sugar cane – which is 50% more costly in terms of energy use as the sugar content in beets is much lower than in cane.

Environmental trade barriers

Here comes a rather tricky issue, namely the seemingly well-intentioned barriers on goods which are considered to have long run negative impact on the environment. Many MDCs have imposed trade barriers on goods such as hardwoods taken from virgin rainforests and crops using certain pesticides. The WTO has come under some intense fire at times since it has generally been rather suspicious of 'green protectionism' and has in most cases not supported green lobby groups in MDCs. The WTO instead takes the position that environmental policies should generally be taken by national governments and not by a trade body.

What muddies the debate waters further is that studies have shown how exporters from the least developed countries are far

more disadvantaged due to environmental trade barriers than other countries. Roughly 20% of the least developed countries' exports are potentially affected by environmental trade barriers, compared to 10% for MDCs.[18] In other words, the very products which many LDCs have a comparative advantage in are the ones causing the most environmental concern.

Non-convertible currencies

The exchange rate examples used in Section 3 had one thing in common; the currencies could be freely exchanged on the foreign exchange market. This is not always the case in developing countries, for a variety of reasons. Many developing countries have fixed exchange rates – i.e. the domestic currency is pegged to another at an **official rate** that is higher than the **market rate** would be, i.e. the domestic currency is overvalued.

Hmmmm, 'excess demand' and 'controls and rationing' for a currency. Does this remind you of something? Yes, parallel market activity will rise – in this case a parallel rate (black market rate) for foreign currency will arise.

_____

18 See http://www.wto.org/trade_resources/quotes/environment/environement.htm
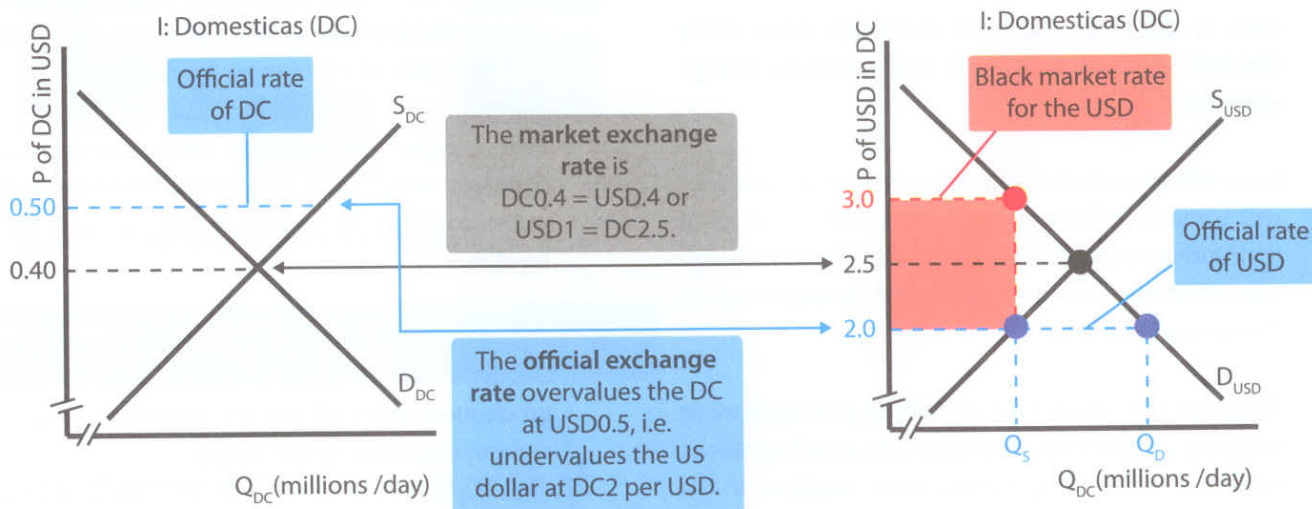
Figure 84.4 Overvalued exchange rate

Diagram II in Figure 84.4 illustrates the outcome of an overvalued exchange rate in a fictitious developing country, using Domesticas (DC) as the domestic currency. (NOTE: diagram I is simply to help you in identifying how a currency is overvalued, yet it is diagram II – the market for US dollars – that illustrates the issue of a parallel exchange rate.) The market rate for the Domestica, diagram I, would be at 40 US cents to the Domestica but the rate has been officially set at a higher exchange rate of 50 US cents per Domestica. Diagram II – for the US dollar – shows how the official exchange rate has set the

price of the dollar too low; at a rate of 2 Domesticas rather that the free market rate of 2.5 Domesticas to the US dollar.

In other words, the official exchange rate overvaluation of the domestic currency has the same effect as a _maximum price_ on foreign currency. Since such a country will not have the foreign reserves necessary to buy up excess Domesticas in order to establish the rate at a higher level, there is _excess demand_ of US dollars ($Q_s \Leftrightarrow Q_D$) and a potential parallel market for the US dollar (shown by the blue rectangle) resulting in a _black market_

595

exchange rate of DC3 to the US dollar. While increasingly uncommon, many countries dealt with this overvaluation of the domestic currency by imposing exchange controls and limits on foreign exchange.[19]

Now, do you think that international foreign currency traders will accept such market shenanigan (= trickery, nonsense, foolery) activities? Not a chance. The foreign currency market won't touch such currencies with a 5 metre barge pole. Even countries which are not *intentionally* trying to keep the domestic currency overvalued often have limited tradability on the foreign exchange markets, as foreign countries will not be prone to trade with currencies for which there are parallel markets. For example, try trading in your birr (Ethiopia), kip (Laos) or gopik (Azerbaijan) at a foreign exchange office in the EU. These currencies will in most cases not be accepted – or even forbidden to be used – outside the country of origin and are therefore considered **non-convertible**.

There are a number of side-effects for a country whose currency is not accepted in other countries, none of them good:

- Currencies facilitate trade, investment and all the other possible economic exchanges going on between countries. An LDC with a non-convertible currency adds to the risks and costs of foreign firms doing business and will *limit trade* and incoming foreign investment.

- Exporters in countries with non-convertible currencies will commonly have to exchange a portion or all of their foreign export earnings – at the official exchange rate. What this amounts to is in effect an *export tax* on domestic producers.

- Black markets for currency will *distort prices* and waste resources as time and effort goes into avoiding official rates and withholding monies from taxation. It has also been amply demonstrated that a fixed exchange rate in the hands of a corrupt and autocratic regime gives many opportunities for those in power to line their own pockets: the dictator simply forces traders to hand over valuable foreign currencies in exchange for highly overvalued domestic currency. (I simply *must* point out that such rulers will have a printing press in the basement!)

HL extensions

How long-term changes in the terms of trade create development barriers in LDCs

Chapter 76 looked at how the *terms of trade* are dependent on the price of exports and the price of imports. Looking at how the price of primary goods has fallen continuously for many years, it is little wonder that the terms of trade for primary goods exporters has deteriorated also. In fact, the terms of trade for primary goods producers have been falling since the 1880s.[20] Adjusting prices to inflation yields even worse figures; the Food and Agricultural Organisation (FAO) estimates that the *real* price of agricultural goods fell by 50% between 1980 and 1999.[21] Historical evidence points to a continuous fall in primary goods prices relative to secondary goods which means that primary goods exporters have seen their terms of trade steadily decline.

A good many developing countries are dependant on three or four primary goods for the better part of export revenue and the results of falling terms of trade for developing countries are almost always negative:

- Higher costs of **debt servicing** as a greater quantity of exports are necessary to earn a given amount of foreign currency with which to repay foreign debt.

- Falling export revenue can cause **current account deficits** – which have sometimes forced countries to increase borrowing – which in turn increases the **debt burden**.

- Deteriorating terms of trade **reduce much-needed imports** such as capital, intermediate products in production, and fuel. All are needed to industrialise and increase value-added output.

- When the market for legal crops falls through the floor, **illegal crops** such as coca become very attractive to poor farmers. This tragic example of producer substitutes can be seen in several South American countries.

- When developing countries' terms of trade fall continuously, there is an incentive to make up for this by increasing output and **exporting more**. The

---

19 In some extreme cases it is illegal for citizens to possess foreign currency, thereby forcing traders to hand over foreign hard currencies at a low rate.

20 *Cypher* and *Dietz*, page 86 and 180

21 FAO at http://www.fao.org/DOCREP/005/Y3733E/y3733e0d.htm

long run effect of this – other than having to pay more for any given amount of imports – is that supply of the commodity increases and helps lower prices further. Between 1986 and 1999, non-fuel exporters of commodities saw the average price of exports fall by 12%, export volume increase by 43% – while the amount of imports that they could buy for total export revenue (i.e. price ´volume) increased by only 2%. In other words, the 2% increase in imports costs 43% more in exports.[22]

Another cycle; terms of trade, Bop and exchange rates

Developing countries dependent on primary goods for export revenue face the possibility of a 'loop' arising from falling commodity prices, illustrated in figure 84.5. A fall in the price of export commodities leads to a deterioration in the terms of trade → low PED for exports → export revenue falls → current account worsens → exchange rate depreciates → fall in the price of exports…
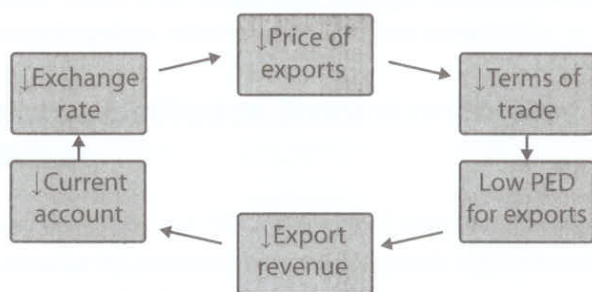
Figure 84.5 Adverse cycle of falling terms of trade

POP QUIZ 84.1

Trade Barriers to Development

1. Prices of commodities continue to fall. Explain why using price elasticities.

2. Why is there an incentive for developed countries to have differentiated tariffs on goods with different levels of value-added?

3. Why might a developing country in fact have two exchange rates?

4. How does a non-convertible currency create a barrier for development?

5. Exporting a narrow range of goods increases a country's vulnerability. Explain.

6. HL: What are the consequences of deteriorating terms of trade for developing countries?

7. HL: Why have the terms of trade moved against a number of developing countries for decades?

---

22 *Least developed countries report*, UNCTAD 2002, page 140

Summary & revision

1. Many of the poorest developing countries are highly dependent on a **narrow range of export commodities** for the brunt of export revenues.

2. **Common development barriers** for developing nations are to be found in international trade: narrow range of export commodities; price fluctuations for exports; and lack of access to international markets.

3. The 'trap' of export commodities derives from the **difficulties of diversifying** into other industries and moving up the value-added chain. Tariff escalation, lack of market knowledge, scant access to technology and low levels of education contribute to this trap.

4. **Price fluctuations and falling prices** over time have characterised primary goods markets for decades. This is primarily due to low PED and yED for primary goods together with supply outstripping demand over many years.

5. LDCs often have difficulty in accessing international markets due to:

   a. Tariff escalation – MDC tariffs can be higher on processed or finished goods than on basic commodities with low value-added

   b. Agricultural subsidies in MDCs seriously skew agricultural prices in favour of MDCs and destroy markets in developing countries

   c. Environmental trade barriers hit LDCs far more than MDCs

   d. Non-convertible currencies limit trade and increase costs of doing international business in LDCs

6. **Worsening terms of trade** over many years has been highly detrimental to LDCs – increased debt servicing, falling export revenues, reduction of imports, and switching to illegal but more lucrative crops such as coca.

85. Trade Strategies I – Import Substitution and Export Promotion

Key concepts:

- Import substitution
- Export promotion

Chapters 85 to 88 look at growth/development strategies from a standpoint of international transactions and organisations. For some six decades the debate has raged between politicians and economists as to which strategy is the most successful in terms of growth/development. Quite frankly, it is impossible to state that *one* particular strategy works better than others since there are large differences between the countries involved and also since the relative success or failure of any given strategy is so heavily dependent on many other variables. There is no 'single best' strategy, but rather a range of connected options which may or may not work.[1]

Import substitution (inward-orientated policies)

During the 1950s and '60s a number of countries attempted to industrialise by substituting imported goods with domestically produced goods. The basic strategy of **import-substitution** is to implement barriers to imports – e.g. tariffs – while perhaps also encouraging domestic producers with subsidies. (This of course requires that there are government funds available.) The rationale is that increased demand for domestic goods will move domestic industries along a learning curve so that they can ultimately compete on equal ground with foreign firms.

Diagrams I and II in Figure 85.1 show the range of possibilities for a developing country implementing import restrictions (or protectionist policies) as a step in inward oriented policies. **Diagram I** illustrates how a **tariff** can reduce imports and

increase domestic output. Domestic output is initially zero and imports are Q_{D_0}; a tariff of $S_{world+tariff}$ will lower imports to $Q_S \Leftrightarrow Q_{D_1}$. Domestic production increases accordingly to Q_S.
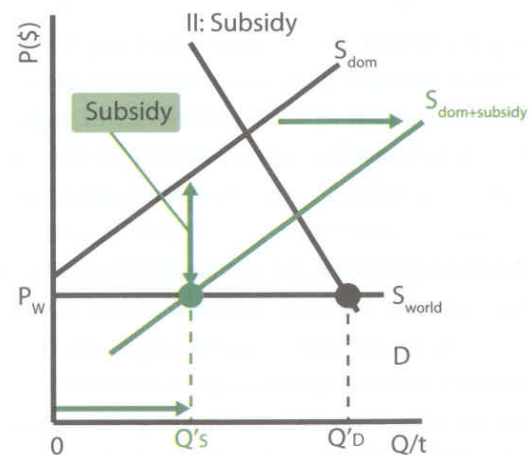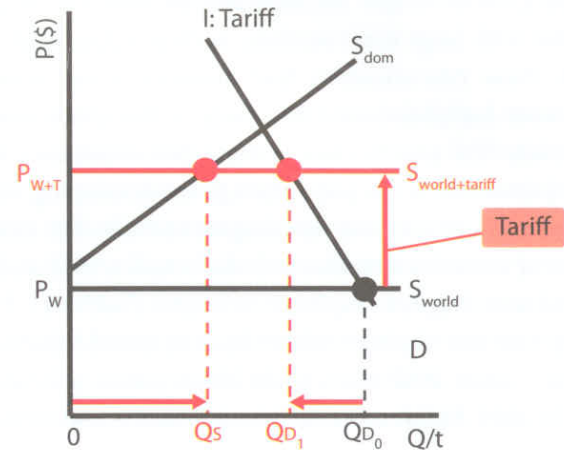
Figure 85.1 Tariffs and subsidies used in import-substitution policies

It is also possible that tariff protection is mixed with a degree of domestic subsidies. Diagram II in Figure 85.1 illustrates how a subsidy increases domestic output from zero to Q'_S. Note that this solution results in the same increase in domestic output as

---

1 There is also an unfortunate propensity towards development 'fads' which come and go; *structural change models* in the 1950s; *inward orientation* in the 1960s, *free trade market liberalism* of the 1980s, and possibly *sustainable development* of the 1990s. The latest catch-phrases from the development Mod Squad of the late 2000s include *capacity building* and *good institutions*. Maybe someday we will get to a development fashion show where the models are wearing *practical, time-tested* and *functional* clothing. While I'm at it, I want a Lamborghini… and one of the models driving it.

the tariff rate in **diagram** I, but that the increase in domestic production due to a subsidy does not increase the market price.

Whatever protectionist mix government ultimately decides on, the aim of what in effect is a *supply-side policy* is to increase output to the point where benefits of scale arise, i.e. where unit costs fall due to larger production scale. This HL concept is illustrated in Figure 85.2, as a movement along the long run average cost curve (LRAC). When the domestic industry becomes competitive at an international level, point C, tariffs and other barriers can be successively removed as the industry can now stand on its own feet.

As the outward oriented industrial development strategy is often based on *infant industry* trade barriers allowing firms to achieve scale economies behind protectionist walls, inward oriented policies might be expected to function better in countries with *large inner markets*, such as China and India. Indeed, these two countries had inward-oriented strategies for decades but abandoned this path in the 1980s and '90s respectively. The growth rates of these two countries – which make up over half of the population in the developing world – escalated considerably after opening up to trade. For example, 30 years of inward orientation in India between 1950 and 1980 rendered an average per capita growth rate of around 1.5% per year – a rate that has been beaten by a factor of 3 during *any single year* since 1992 when trade liberalisation policies took effect. By 2004, India's economy had doubled in real terms.[2]

Failure of import substitution

There is a great deal of controversy concerning the degree to which outward oriented policies are the main source of the growth differences. A frequently referred-to example is that, during the 1950s, Korea and India had very similar per capita incomes yet in the following two decades Korea grew at a far faster pace – and the other three Asian Tigers (Taiwan, Singapore and Hong Kong) grew even faster than Korea. The conclusion is that the outward oriented policies of the Asian Tigers have been far more efficient in terms of growth than India's strongly inward oriented policies up to the 1990s. Yet some economists attribute considerable weight to the protectionist and government interventionist policies of the Tigers.[3] There was also a strong *political* element in many developing countries' choice between inward and outward-orientation; many developing countries wished to prove their independence from western – often colonial – powers.

Whatever the actual cause of the growth differences between outward oriented countries and inward oriented countries, the fact remains that most countries that pursued the latter *failed to achieve substantial growth or development*:[4]

- The main failing was probably due to *lack of competition*. Domestic firms produced high-priced and shoddy goods – which did little for the domestic economy and even less for potential exports. Nor have governments been particularly adept at choosing the right industries to protect and subsidise. India's 'National Champions' of domestic industry have been competitive failures in almost every case, producing shoddy goods at inflated prices whilst enjoying the 'quiet life' of a monopoly protected by cronies in government.

- Import restrictions were often far lower on capital goods, since these were needed in domestic production. This limited the creation of employment and thus income which could fuel consumption and saving/investment. This also countermands the entire foundation of *comparative advantage*, as goods which are labour intensive will be disadvantaged when use of capital is favoured by government policy. Additionally, capital is often imported, as are components for upkeep/renovation/repairs and labour skills required for assembly and service – all of which will burden the current account in the balance of payments.

- Another aspect of not focusing on comparative advantage is simply that *governments are not good at picking 'winners'*, i.e. firms and industries that can become successful in competing with foreign producers. India most assuredly did not have a comparative advantage in producing cars but levied a 800% tax on imported cars in order to produce the Ambassador car which remained virtually unchanged between 1950 and 2003.

- Another factor was pure outright *corruption*; protected economies will unerringly create smuggling, tax evasion and bribery.

- A great deal of industrial output is used as inputs in other industries. By levying tariffs on imports needed for production, costs will rise for domestic producers; this is the *forward linkage effect*. The main import-

---

2   http://in.rediff.com/money/2004/jan/24wef1.htm

3   See for example Todaro & Smith, pages 620 – 653

4   See for example *Once more into the breach: economic growth and global integration*; Center for Global Development, Working paper no 34, December 2002, page 33.

substitution countries in Latin America had huge average tariffs on both imported consumer and capital goods; 131% in Argentina, 168% in Brazil, 138% in Chile, 112% in Colombia and 61% in Mexico. In the latter case, Mexico, the *effective* rate of import taxes was 671% on fertilizers and insecticides, 206% on pharmaceuticals and 102% on cars.[5]

All the above contribute to slow growth and low proportion of exports relative to GDP. For example, Brazil has had a notable policy of promoting domestic production by curtailing imports, and has seen how exports have remained pretty much unchanged (around 7% of GDP) between 1970 and 1998.[6] However, it is worth noting that the line between an import-substitution and export led strategy is very fine when you consider that a good many outward oriented countries implemented infant industry trade barriers and export subsidies. What stands out is that countries which focused on import-substitution as a **main** strategy of growth in most cases opted out in the 1980s and 1990s, in order to put emphasis on markedly outward oriented policies.

Hindustan Motors, Ambassador car – guess the year![7]

Export promotion (outward orientated strategy)

The strategy which a number of developing countries chose in the 1960s was to concentrate on producing goods specifically aimed at export markets. Generally, such a strategy involves focusing on industrialisation and opening the economy by

5   *Import substitution industrialisation* at www.colby.edu/economics/faculty, page 59

6   *Irwin*, pages 76 – 78. It should be noted that Brazil in fact has had a markedly outward oriented strategy for *primary goods* – which have fallen drastically in price and thus not enabled an increase in export revenue.

7   This is the 2001 model.

moving towards free (-r) trade and free (-r) capital flows from abroad. Countries such as South Korea, Singapore, Hong Kong, Taiwan (the 'Asian Tiger' economies) and Malaysia have shown phenomenal growth rates since the 1960s – over 7% on average during the past 40 years compared to 2-3% for the US and Europe – most of which is attributable to strong export-led/outward-oriented policies. India and China have had growth rates of between 6 and 9% during the 1990s due to increased liberalisation in trade and investment flows.

Characteristics of outward orientated strategy

These **newly industrialised countries**, NICs, are of course very diverse and have great differences in trade policies and levels of government intervention. Yet there are a number of broad characteristics of countries which have adopted an outward strategy:

- In order for increased trade to take place, **trade barriers are lowered** on goods and services. Yet one should take care in claiming that all barriers for all goods have been lowered in outward oriented countries. In fact, several of the Tiger Economies have been rather protectionist initially, targeting certain central industries for **infant industry protection**. For example, during a period of extreme growth in the mid-1980s, average protection rates in the Tiger Economies was still 24%.[8]

- **Capital markets** are opened and regulations on capital flows are relaxed. This facilitates inward portfolio and direct investment by foreign firms – often multinationals. All of the NICs have been open to inward investment flows and attracted a great deal of direct investment from foreign multinational companies.

- In most cases there has been a focus on goods which are **labour intensive** in manufacturing, such as shoes, leather goods and apparel, i.e. the economy has concentrated on goods for which there is a competitive advantage.

- **Government involvement** has been notably strong – contrary to many beliefs on the 'free market' content of

8   Krugman & Obstfeld, page 268. Many economic historians also point out that the 'beacons' of modern free market thinking – USA and UK for example – all had long periods of fierce protectionism in the 1800s which today would be called import-substitution policies.

the NICs. (Do not confuse *"laissez faire"* with market-based policies. Korea, Singapore, Taiwan and Japan have been strongly market-based economies but with a notoriously strong element of government planning.) A variety of government policies have supported investment and industries: production and export subsidies to key industries; land grants; soft loans; and generous tax write-offs for profits ploughed back into investment. Thus it should be noted that while there is indeed correlation between trade and economic growth, there is wide disagreement as to which came first. Several economists conclude that the success of Asian outward oriented countries in trade might well be primarily the result of **interventionist polices** and strong **saving, investment** and focus on **education**.

There are numerous other export-promoting methods which developing countries can apply. For example, the currency can be kept artificially low to promote exports and deflect imports; tariffs on much needed factor imports can be removed; and the establishment of 'free trade zones' and industrial parks for export firms.

Reasons for outward orientated strategy

A glance at the size of Hong Kong, Singapore and Taiwan provides a clue as to why an outward oriented strategy of growth was adopted; small populations and therefore *small domestic markets*. Yet even Korea, with its 46 million people, had a mere fraction of the potential markets available when the US and Europe were included. From Adam Smith onward, economists have pointed out that market size matters. Here are a few notable points on how an outward oriented strategy can benefit a country:

- **Increase in export revenue:** Increased exports as seen in the NICs have had an immense impact on the domestic economy. Exports grew at an average rate of between 8% and 20% between the 1960s and 2000 in the NICs, and most of the exports are manufactures.[9] This has been a tremendous boon to the domestic economies – for example, the proportion of exports to GDP in Korea went from just 2.4% in 1962 to 42% in 1999.[10]

---

9   *World Development Indicators*, World Bank 2002
10 *Legrain*, page 69

- **Benefits of scale:** When the size of the market increases due to open markets, then it becomes possible for small countries to establish economically viable plant sizes to compete internationally. Longer production runs, better use of resources and the spreading of fixed costs are examples of how the average costs fall for firms increasing the scale of operation.

  o Figure 85.2 shows how average costs in an industry continuously fall as scale benefits rise. As the industry *output increases faster than additional costs*, the average cost of each unit falls. This renders the long run average cost (LRAC) curve shown in the diagram. A developing country at point A will not be able to compete with international – foreign – competitors, so it must seek to move towards C in order to have an export competitive price. In order to initially move towards point B, the country might subsidise its industries and/or impose trade barriers on imports. This is the classic *infant industry argument* in support of an outward oriented policy.
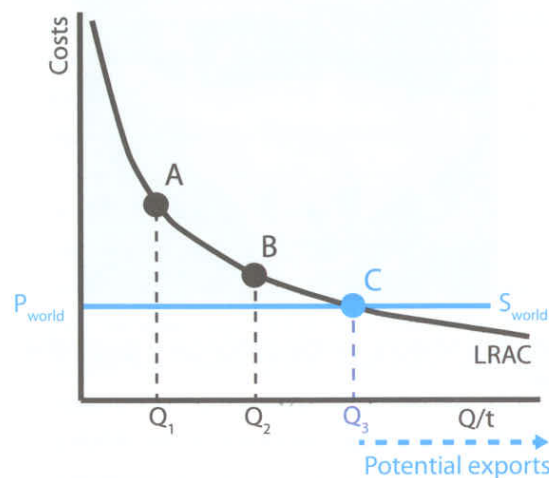
Figure 85.2 Falling long run averages costs (LRAC) in an industry

- **Increased competition:** When a country competes internationally, there will be far more incentive for domestic suppliers to increase efficiency and product quality goods, rendering dynamic gains over time.

- **Additional resources:** An outward oriented country stands to benefit not only from increased exports but from increased imports, i.e. raw material, components and intermediary goods needed in production. About two thirds of all manufacturing output is sold as inputs

to other firms, which means that access to global networks of suppliers lowers production costs.

- **Technology transfers:** In addition to this, there is technological transfer, both in the form of capital and in the form of knowledge and experience imparted by foreign firms' direct investment inflows.

Evaluation of outward orientated strategies

The relative success or failure of outward oriented policies is part of the globalisation issue and thus subject to a heated debate. I would, however, dare say that the majority of economists would judge outward-orientation a success on the whole. For example, Ghana had roughly the same GDP per capita as South Korea in the 1960s, and by 1996 Korea had joined the 'rich man's club', the OECD, having increased per capita income by a factor of 9.[11] The figures for the other Asian Tigers are similar and relative late-comers have also shows strong growth: China had 8 – 10% growth rates during most of the 1980s and 1990s and India's growth has been around 7%. Both countries have seen this growth after large-scale reforms to increase openness and remove polices which hampered trade; China in the early 1980s and India some 10 years later.

It would be wonderful if I could just stop right here, but there is, of course, another side. A number of issues have been raised in opposition to outward oriented policies, many of which you will have studied in Section 3: *low wage and environmental policies* might be intentionally set by developing countries in order to attract foreign investment, leading to exploitation of labour and environmental degradation; dependency on multinational companies could lead to the empowerment of foreign businesses over *domestic policies*; the domestic economy for a strongly export oriented country would be *dependent* on international business cycles; and *speculative inflows* can have devastating effects on the economy such as the 1997 Asian Crisis (see Case Study in Chapter 72).

Yet perhaps the main problem is simply that developing countries opening up to trade will initially be far from competitive, and opening their economies along the lines of free trade will *threaten domestic output* and employment as very efficient foreign suppliers enter the market. For this reason, many economists consider that there is merit in limited – infant industry – trade barriers. There is also scant supporting evidence that what works in one country will inevitably work in another.

Once again, the relative success or failure of a development strategy does not stand alone but depends on all the other factors which enable development. Korea and Singapore were remarkably successful because the tax revenues from growth were to a large extent immediately poured back into *education*. In addition to this were high saving rates which contributed to investment; a relatively high rate of initial literacy; and political stability – all of which were noticeably lacking in a number of other developing countries.

A 30 year summary from UNCTAD

The United Nations Conference on Trade and Development, UNCTAD, publishes a yearly report called the Trade and Development Report. In the 30th anniversary edition in 2011, 30 years of development thinking is summarised and some effort is devoted to inward and outward orientated strategies. I end this chapter with a few brief points taken from the report.[12]

- A number of Trade and Development Reports (TDRs) have made it clear since the 1990s that it is most assuredly *not a case of X or Y* in terms of export orientation or import substitution. These are not mutually exclusive strategies as seen by the successes of East Asian economies where both strategies were in fact integral parts in the long term goals of increasing productivity, investment and growth.

- The successful economies using export promotion strategies had in fact *already built up domestic production capacity in order to replace imports*. The success of Korea simply cannot be copied into the Ghanaian economy – trade policies must into account the level of development, resource endowment and macroeconomic variables.

- Many of the 'early adopters' of *export led growth took the step during the high growth period in the 1960s and 1970s* which enabled them to produce goods for the growing markets in first world countries. Many 'late comers' in the late 1990s focused on labour-intensive low-quality manufactures in similar markets which pushed world prices down.

- And thus; a *sustainable growth strategy* requires greater reliance on domestic demand than has been the case for the past 30 years.

---

11 *Norberg*, page 62

12 TDR 1981 – 2011, pages 47 – 49

1. "An outward-oriented strategy means that a country moves towards free trade." Discuss this statement.

2. Why might an export-oriented strategy appeal to smaller countries?

3. Why was the strategy of import-substitution generally abandoned in most developing countries in the 1980s and '90s?

Summary & revision

1. **Import substitution** (inward-orientated policies) is a strategy aimed at replacing imports with domestically produced goods. Common policies include:

    a. High **barriers to trade** (often using the infant industry argument)

    b. **Government ownership**/control of key industries

    c. **Centralised steering** as to which are the core industries

    d. Focus on **industrialisation**

    e. **Exchange rates fixed** – often together with currency controls

2. Import substitution was **mostly abandoned in the 1980s** due to many shortcomings of the strategy:

    a. **Lack of competition**, high prices and low quality goods

    b. Governments picked industries for **political rather than economic** reasons

    c. **Forward linkage effects** on high-priced imported capital

    d. **Monopolies** evolved with close ties to governments – the 'national champions'

    e. Lengthy red tape and bureaucracy created **corruption**

3. **Export promotion** (outward orientated strategy) focused on producing goods for foreign markets. Commonly:

    a. **Lowering of trade barriers** – but with many NICs using infant industry protectionism

    b. Open and **deregulated capital** markets

    c. Focus on **labour intensive manufacturing** – cheap and simple goods initially and then a shift up the value-added chain

    d. Strong **government involvement in key industries** via subsidies, grants and tax benefits

4. **Reasons** for export promotion include attaining **economies of scale**, importing **capital** and **know-how** from abroad, **export-led growth** and the transfer of **knowledge** and **technology**.

5. The strategy of export promotion has generally been regarded as successful. Many of the Asian countries that adopted this strategy have had far higher growth rates over the past 30 – 40 years than countries with long periods of import substitution.

86. Trade Strategies II – Trade Liberalisation and the WTO

Key concepts:

- Trade liberalisation
- The WTO and developing nations

"Audiences the world over seem to believe that this signifies a set of neoliberal policies that have been imposed on hapless countries by the Washington-based international financial institutions and have led them to crisis and misery. There are people who cannot utter the term without foaming at the mouth."

John Williamson on his coining of the phrase "Washington Consensus"

Trade liberalisation

In 2005 the World Bank estimated that a reduction of LDCs' average tariff rates from 13% to 10% on agricultural produce and 5% on manufactured goods, the gains would be in the area of USD100 billion by 2015 – greater than the gains LDCs would see if developed countries reduced tariffs towards them.[1] The WTO has continuously pointed out that trade liberalisation leads to economic growth, raises labour and environmental standards and decreases child labour.[2]

There are numerous studies from international organisations supporting the above quotes which leads to a certain wonderment as to why 'liberalisation' became such a value-laden term during the 1990s. In fact, it became more of a dirty word for the anti-globalisation crowd which erupted on the scene with violent demonstrations during the late 1990s. Trade liberalisation in its broadest sense means lowering or obliterating any rules or regulations that hinder the free flow of goods and services between countries. Skim back to Chapter 65 and have a quick think about what 'liberalisation' might mean before reading on.

The origin of what most economists would deem the 'modern usage' of the term has its origins in the debilitating and reoccurring economic crises in Latin America during the 1980s.[3] An economist named John Williamson put forward in 1989 what he considered was a fairly straightforward and commonsensical ten point program for Latin American countries needing to put their houses in order. These strongly *pro-market and liberal policies* were endorsed by, amongst others, the US Treasury, the Inter-American Development Bank, the World Bank and the IMF – all of which are headquartered in Washington.

Williamson's list became known as the '**Washington Consensus**' and it outlined the points a developing country (in fact 'Latin American country' originally) should implement as the "…lowest common denominator of policy advice…" The Washington Consensus dealt with domestic macro-stabilisation, openness in trade and allowing market forces to work – in other words, far broader prescriptive medicine than trade only. Pure trade liberalisation policies in blue below:

1. Fiscal discipline (balance the budget, cut down debt)

2. Redirect government expenditure to areas with higher social returns (health care, primary education and infrastructure)

3. Tax reform (to lower marginal rates and broaden the tax base)

4. Interest rate liberalization (unfettered in terms of a fixed exchange rate)

5. A competitive exchange rate (floating rather than fixed)

6. Trade liberalization (lower barriers to trade)

---

1 World Development Report 2005, page 112
2 http://www.wto.org/trade_resources/

3 I cannot recommend the PBS series *Commanding Heights* strongly enough. It is brilliant.

7.   Liberalization of capital/financial account (freer flows of FDI and portfolio investment)

8.   Privatization (selling off government assets)

9.   Deregulation (to abolish barriers to entry and exit)

10.   Secure property rights (rule of law and ownership of land rights)

The Washington Consensus dealt with domestic *macro-stabilisation, openness in trade* and allowing *market forces* to work – in other words, far broader prescriptive medicine than trade only. The Structural Adjustment Programs of the IMF and World Bank (see Chapter 93) have their source in the Consensus and much of the fury directed against the two 'Sister Organisations' in the 1990s is a direct result of what critics deemed was a neo-liberal and conservative line of politics propagated by the IMF and World Bank.

Evaluation of trade liberalisation

The bag of outcomes here is so mixed that I actually wonder if it's worth writing about. The results can apparently be cherry-picked to the extent that it is possible say *whatever* the author wishes. Some of the most famous development economists in the world arrive at what are in fact almost opposing results – and to worsen matters, evaluation outcomes from the 1990s have often been heavily criticised by more complete data ten to fifteen years later. So, rather than a tiresome 'on-the-other-hand ping-pong match' I give you the following six very condensed points:

*   **Trade liberalisation and growth**: India, China and the Asian Tigers experienced phenomenal growth rates during the 1990s while Latin America lagged and Africa in many cases saw contraction. The weakest growth was seen in countries predominantly exporting commodities.[4]

    o   It should be pointed out that both China and South Korea 'disobeyed the rules' of trade liberalisation to a large degree. Both had very strong elements of government control and central planning by either ownership or decree, and both had substantial protectionist policies in place for what were considered strategic industries.

o   **Trade liberalisation and income distribution**: The gap has widened and there is broad agreement on this. In 1965 the G7 (seven richest countries in the world) were 20 times wealthier than the poorest 7 nations. By 1995 this had grown to 39 times. The income share of the richest 20% has risen in virtually all developing countries – in half of the LDCs the richest 20% receive half of GDP.[5] One of the most extreme examples is China where the Gini coefficient has gone from 0.27 in the early 1980s to just under 0.4 some 25 years later – see Figure 86.1.
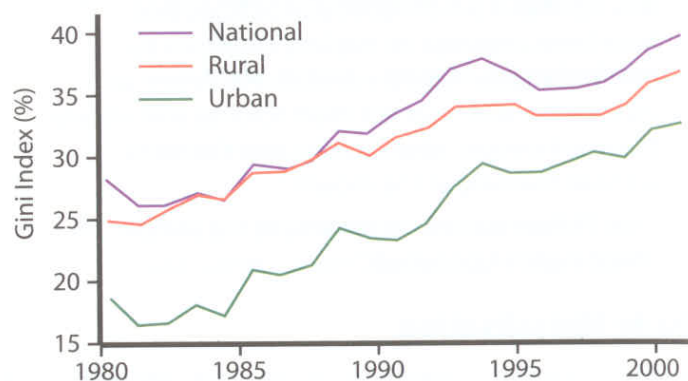
Figure 86.1 Inequality in China over 25 years (Source: http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20634060~pagePK:64165401~piPK:64165026~theSitePK:469382,00.html)

*   **Trade liberalisation and productivity**: One of few areas of agreement; productivity took a leap forward in all 'globalising' countries. Recent research shows that after India liberalised its trade rules in the early 1990s, few of the firms subsequently subjected to competition actually went bankrupt. It is estimated that for every 10% reduction in tariffs that *total* factor productivity increased by 0.5%.[6]

*   **Trade liberalisation and poverty**: This is where – to paraphrase Mark Twain – economic commentators can say "…any damn thing they like…" by playing with the numbers. What is clear is that some 600 million have been taken out of extreme poverty (defined by the World Bank as living on less than USD1.25 per day adjusted for purchasing power) over the period 1980 to 2000 – two thirds of this in China alone.

---

4   IFM   at   http://www.imf.org/external/np/exr/ib/2001/110801.htm

5   GDS 2012, page 54

6   WDR 2005, page 66
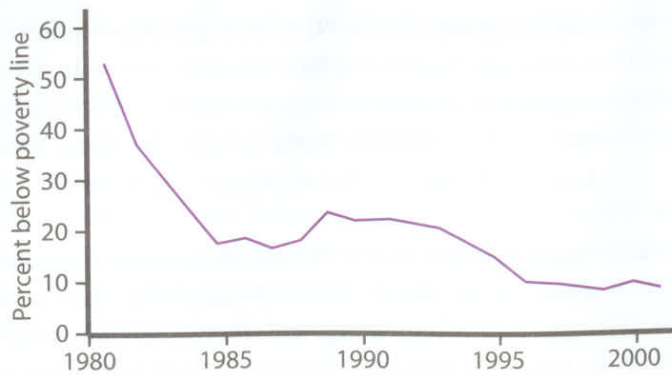
Figure 86.2 Poverty in China over 25 years

Source: http://econ.worldbank.org/WBSITE/
EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK
:20634060~pagePK:64165401~piPK:64165026~th
eSitePK:469382,00.html

Region/Variable - Period	P.C. GDP growth	
	1981 - 95	1996 - 05
East Asia and Pacific (EAP)	6.894	6.355
Eastern Europe and Central Asia (EECA)	−3.434	4.138
Latin America and Caribbean (LAC)	0.140	1.394
Middle East and North Africa (MENA)	0.713	2.309
South Asia (SAS)	3.208	4.143
Sub-Saharan Africa (SSA)	−1.009	1.293

$1.25 standard	1981	1996	2005	1981-1996
EAP	77.67	36.00	16.78	−2.78
EECA	1.67	4.61	3.65	0.20
LAC	12.87	10.94	8.22	−0.13
MENA	7.87	4.10	3.60	−0.25
SAS	59.35	47.05	40.34	−0.82
SSA	53.37	58.78	50.91	0.36
China	84.02	36.37	15.92	−3.18
China (rural)	94.08	49.48	26.11	−2.97
China (urban)	44.48	8.87	1.71	−2.37
India	55.51	49.40	41.64	−0.56
India (Rural)	57.78	52.46	43.83	−0.48
India (Urban)	48.25	40.77	36.16	−0.68

- **Trade liberalisation and agriculture**: One of the clearer results shown is that the agricultural sector has seriously lagged behind in countries that have opened up to trade. While secondary sectors have been able to reap gains of openness and increased exports, large portions of populations dependent on agriculture have seen little benefit. For example, by 2004 the agricultural sector in India employed some 60% of the labour force but accounted for less than 2% of exports.[7]

- **Trade liberalisation and regions**: Compiling data for growth and poverty per region, we return to the issue of the 'mixed bag' of results previously outlined. Figure 86.3 indicates that the 'first mover globalisers' in South and East Asia not only had the highest growth rates but the largest decrease in poverty. Most notably, China has lowered the percentage of the population living in poverty from 84% to just under 16%. India and East Asia/Pacific shows a similar remarkable decline in absolute poverty levels. Depressingly unremarkable is Sub-Saharan Africa which saw poverty levels *climb* during the 1980s and 1990s before falling somewhat below 1981 levels by 2005.

Figure 86.3 Growth and poverty per region 1981 – 2005 (Source: OECD background paper for Global Development Outlook 2010, 'Growth, inequality and poverty...', tables 1 and 2, March 2010)

The WTO and developing nations

"...what we actually achieved at Doha was a significant breakthrough".[8] Mike Moore, former Director-General of the WTO (1999 – 2002)

The WTO (Chapter 63) has increasingly focused on some of the key development issues facing the world today – primarily the unresolved issue of rich world trade barriers for products which make up the brunt of the poorest nations exports such as textiles and agricultural goods. The WTO is the only true

7 http://archive.unctad.org/Templates/webflyer.asp?docid=4919 &intItemID=2807&lang=1

8 *Moore*, page 175

DEVELOPMENT

worldwide trade organisation, and in sketching an outline of the 'successes and failures' of the WTO, I came upon a somewhat surprising pattern, namely that economists and free-trade advocates generally regard the WTO's successes as outweighing the failures while 'global opinion' voiced through any number of anti-globalisation non-government organisations (NGOs) take a predominantly opposite stance. I fully realise the danger in sweeping generalisations like this and I use the two captions below somewhat tongue-in-cheek in trying to represent the different perspectives of the WTO record.

The 'free-trade economist's' view

Supporters of the WTO point out that 600 bureaucrats, two thirds working as translators, with an annual budget one fourth of the World Wide Fund for nature (WWF – one of many vocal NGO critics) do not quite qualify as an evil conspiracy attempting to rule the world. Nor has any reason been put forward why 2.5 million members of Greenpeace should have anywhere near the influence on the WTO as the Indian trade delegate elected by 600 million voters in India. The success of the WTO is primarily in its capacity as a **democratic forum for discussion and a medium of same-rules** for both large and small countries.

- **Representation:** The WTO is far more accountable to its members than most NGOs, and is in fact more *democratic* than most of the NGOs, since every member has an equal vote and all decisions are made by consensus, giving each country a **veto** on decisions. Countries apply for membership and all agreements are undersigned by duly elected representatives of each country.

- **Pro-poor:** The WTO has increasingly put forward developing country issues - the Doha Development Agenda is a case in point. *Rules and representation* will allow developing countries to negotiate more equitable terms of trade and thus join in the economic success evidenced in increasingly free-market countries such as China and India.

- **Equality:** The WTO does not pander to individual lobby efforts – that is done at national level. Thus the WTO acts to *protect smaller members* from strong domestic groups such as US steel and European farmers.

- **Developing countries:** The WTO now includes a good many developing countries – over three quarters of

the members in fact. This provides poor countries with a solid *negotiating base* for rectifying some of the highly unjust trade barriers erected by rich countries. Developing countries stand to gain far more from lower trade barriers than developed countries.

- **Trade barriers:** The WTO has been *largely successful*; tariffs have come down considerably; non-tariff barriers are decreasing and becoming increasingly outlawed. The result is that world trade has increased 20 fold and living standards and output are higher than ever. The Uruguay Round added an estimated 0.74% to real incomes on a global basis.[9]

However, there are a number of points where the WTO has failed, according to the 'free-trade economist's' view:

- It allows **Regional Trade Agreements (RTAs) such as NAFTA and the EU.** Such regional exceptions to the WTO policies of non-discrimination and equal treatment for all, undermines global efforts to create a level playing field in trade.

- **Subsidies and export subsidies** have changed little during the time since the Uruguay Round was completed. The main issue of rich developing countries' enormous subsidies to domestic farmers still remains unsolved.

- There are still noticeably higher trade barriers on goods which are the most important to developing countries, i.e. **agricultural goods and textiles.** The continued breakdowns of negotiations in the Doha Round are a case in point.

The 'grass-roots anti-globalisation' view

"The World Trade Organisation is carrying out a slow-motion coup d'état over democratic governance worldwide".[10]

The detractors of the WTO often take a standpoint where the organisation is a non-democratic and non-transparent tool of rich countries and multinational companies, where poor countries and the environment are victims.

- **Undemocratic power:** The WTO transfers power from national assemblies to an organisation which

---

9 *Salvatore*, page 286
10 *Wallach & Sforza*, The WTO, *Five Years of Reasons to Resist Corporate Globalization*, page 13

is squarely in the hands of *unaccountable corporate interests* aimed at increasing profit via increased trade. Since WTO rules are binding for signatory countries, national laws will have to adapt to WTO regulations.

- **Race to the bottom:** Freer *mobility of capital* will encourage poor countries to attract foreign investors by lowering standards on labour and environment. Multinational companies will 'race to find the countries with the lowest bar' in terms of labour and environmental standards – in effect 'exporting' pollution and harmful labour practices.

- **Investment and intellectual property:** Allowing multinationals to freely allocate investment around the world further weakens the position of poor countries, as they increasingly become *dependent on foreign firms*. Granting global intellectual property rights to big business around the world (see TRIPS above) in pharmaceuticals and crop varieties would cause thousands of deaths in developing nations, as poor people could not afford monopoly prices on medicines and seed grains.

    o *Quote*: "More stringent protection for patents will increase the costs of technology transfer. Developing countries will lose approximately $40bn a year in the form of increased license payments to Northern-based TNCs, with the USA capturing around one-half of the total. Behind the complex arguments about intellectual-property rights, the TRIPs agreement is an act of institutionalized fraud, sanctioned by WTO rules." Read more at http://www.maketradefair.com.

- **Environment:** The WTO puts *free trade over environment*, by allowing countries to file complaints against rules devised to protect the environment and plant/animal species. For example, when the US imposed strict rules on clean gasoline, Venezuela and Brazil complained that this more stringent regulation in effect was a trade barrier to their exports of gasoline. The WTO ruled in favour of Venezuela/Brazil and the US softened its regulations.

- **Labour standards:** The WTO has not implemented safeguards on labour standards, allowing multinational companies the ability to *exploit labour* in low labour countries.

- **Public health:** In insisting that countries allow free trade, the WTO forces countries to relax regulations not on public health which could be considered trade barriers, allowing *imports of harmful goods*. For example, the EU ban on American beef containing trace elements of artificial hormones was deemed illegal by the WTO. (The EU has thus far ignored the WTO ruling.)

Both sides have a number of weighty arguments and it is virtually impossible to assess the final outcome in terms of net benefits. Most of the arguments put forward are non-economic arguments and are properly dealt with in the arena of democratic institutions. The issues which can be dealt with in economics are often very difficult to summarise, such as the problem of when a ban on a certain good due to health/safety reasons is a valid concern or simply hidden protectionism. I should mention that a good many commentators feel that some issues, such as labour standards and child labour, should not be dealt with by the WTO.

Summary & revision

1. **Trade liberalisation** during the 1980s and 1990s was encapsulated by the so-called 'Washington Consensus' – a 10 point liberalisation program aimed at developing countries. The main points are macro-stabilisation, openness in trade and free market reforms.

2. The main points specific to **trade liberalisation** are:

   a. Freely *floating exchange* rates

   b. Lower *barriers to trade*

   c. Allow *freer flows of FDI* and portfolio investment

3. **Results of trade liberalisation** are highly debated and politically contentious, yet a few points can be made reasonably risk-free:

   a. *Growth in liberalised countries* has outpaced growth in inward looking countries for the most part

   b. Liberalisation of trade is linked to *increased income inequality* within newly-liberalised countries

   c. Increased openness to trade *increases factor productivity*

   d. There is ample evidence of *strong correlation between poverty reduction and trade* liberalisation

   e. In countries with a high proportion of the population involved in *agriculture*, trade liberalisation has had *minimal impact and worsened income distributions* between urban and rural areas.

   f. There is a *wide divergence of growth and poverty reduction* between regions. East and South East Asian countries have seen far higher growth rates and poverty reduction than Latin America and Sub-Saharan Africa.

4. The WTO has received both positive and negative critiques from economists:

   a. *Positive*: The WTO is highly representative and democratic, pro-poor, egalitarian and has managed to lower (visible) trade barriers significantly over 50 years.

   b. *Negative*: The WTO allows discriminator RTAs and has not managed to deal with the issue of rich country subsidies for goods such as agricultural goods and textiles in which LDCs have a comparative advantage.

5. **Anti-globalisation proponents** put forward that WTO power is stacked in favour of powerful nations and MNCs. This leads to dependency and a 'race to the bottom'.

87. Trade strategies III – regional trade agreements (RTAs)

Key concepts:

- Bilateral preferential trade agreements (PTAs)
- Regional PTAs
- Evaluation of PTAs

"When a nation "exempt[s] the good of one country from duties to which it subjects those of all others ... the merchants and manufacturers of the country whose commerce is so favoured must necessarily derive great advantage".

Adam Smith

Regional trade agreements (RTAs), outlined in Chapter 74, have proliferated greatly in the past years. It is quite possible that some of the reason lies with the fact that the WTO Uruguay round was completed in 1994 but the Doha Development round which began in 2001 simply has not come to fruition. There are now close to 300 PTAs registered with the WTO.[1] Figure 87.1 shows the marked increase in PTAs from the early 1990s onwards.
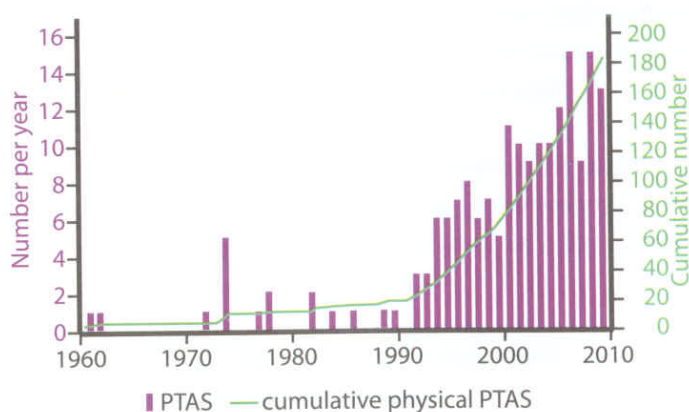
Figure 87.1 PTAs notified to GATT/WTO and in force, by year of entry into force 1959-2009 (Source: 'Preferential Trade Agreement Policies for Development, World Bank publication, 2011, page 41)

Bilateral preferential trade agreements (PTAs)

Recent data shows that PTAs are increasingly *bilateral*, i.e. between two countries that are not necessarily regionally linked. Let us – out of what might be called a 'spaghetti bowl' of overlapping examples – briefly mention Thailand which has bilateral PTAs with Laos, China and Australia. There is also an agreement with New Zealand where all tariffs between the countries will eliminated by 2025.

Regional PTAs

A regional PTA consists of several countries agreeing on eliminating or decreasing barriers to trade within the member zone. Continuing with the example above, Thailand belongs to the Association of South East Asian Nations (ASEAN) comprised of 10 South East Asian countries.[2]

Evaluation of preferential trade agreements

Keeping firmly in mind that there are considerable differences between the scope and depth of PTAs and the member countries, the World Bank considers the following points in its 2011 assessment of PTAs and development:[3]

---

1 *Preferential Trade Agreement Policies for Development*, World Bank publication 2011, page xi

2 ASEAN consists of Brunei, Cambodia, Indonesia, Laos, Malaysia, Burma, Philippines, Singapore, Thailand and Vietnam

3 *Preferential Trade Agreement Policies for Development*, World Bank publication, 2011

- PTAs are generally conducive to trade and there is ample evidence of increased trade within PTAs. It bears noting that the correlation between intra-community trade and export growth is much higher for REAs such as the EU (regression coefficient of 0.98) and ASEAN (0.88) but considerably lower for smaller REAs such as SAFTA (0.33) and CEFTA (0.31).[4] (RTA handbook page 61)

- They create positive dynamic effects over time– notably increased technology transferral and FDI increase LR potential growth. For HL: there is solid evidence that the trade creation effect outweighs the trade diversion effect in *regional* multilateral PTAs.

- They have proven effects on income and poverty levels – the World Bank points out that no low income country has managed to grow and reduce poverty without regional trade.

- And yes, there are a few 'howevers'. The most commonly put forward is that PTAs hamper global openness and multilateral trade. In effect, they can create blocks to outsiders and discriminate against non-members. Having said that, the World Bank states simply that there is little evidence that 'regionalism' is "…overwhelmingly bad for the multilateral trade system, as some had feared."[5]

- Market failure in terms of monopolies and cartels is also a possibility in PTAs as larger and more powerful firms gain access to new markets. The World Bank points out that having solid rules and regulations in the trade agreement can, to a large extent, avoid this problem.

Summary & revision

1. **Preferential trade agreements** (PTAs) are designed to lower or do away with trade barriers between the signatory countries. PTAs can be *bilateral* (between two countries) and *regional* (between a number of countries in a region).

2. The **number of PTAs** in the world have increased rapidly since 1990, numbering close to 300 in 2011. A possible explanation is that multilateral agreements – WTO rounds – have stalled since 2001.

3. There is **evidence of positive effects** such as *increased trade, income* and *poverty reduction* in PTAs and positive long run dynamic effects such as the *spread of technology and knowledge*.

4. **PTAs have been criticised** for blocking multilateral trade and creating barriers for non-members. Regional PTAs can also lead to market failure in the form of cartels and monopolies.

---

4 IBID, page 61. SAFTA is the South Asia Free Trade Area; CEFTA is the Central European Free Trade Agreement.

5 Ibid, page 6

88. Trade strategies IV – Diversification of Exports

Key concepts:

- Diversification of exports
- Strategies to diversify

As shown in Chapter 84, one of the many problems facing LDCs is that so many of them are exporting goods within a very narrow range of commodities. Such goods are commonly highly homogeneous and have low value-added – most of the value-added comes from importing countries where the raw materials are processed into secondary goods. In short; the profit goes abroad rather than staying in the country of origin. The key question here is how LDCs can move 'up and out' – up the value-added chain and out into other industries.

Diversification of exports

The World Bank's International Trade Department has had an on-going project on how developing countries might stimulate export diversification. It has been clearly identified that increased diversity in exports is linked to higher growth rates – see Figure 88.1. The core outlook is that increased export diversification has three key roles in development:

1. It reduces the impact on exporters of large price fluctuations and international contractions in the business cycle. This reduces reliance on a few low-price goods for export revenues.

2. Increasing the at-home value-added component increases income from exports.

3. Expanding production to non-traditional areas adds to expertise and enhances overall human capital which in turn increases long run potential supply.
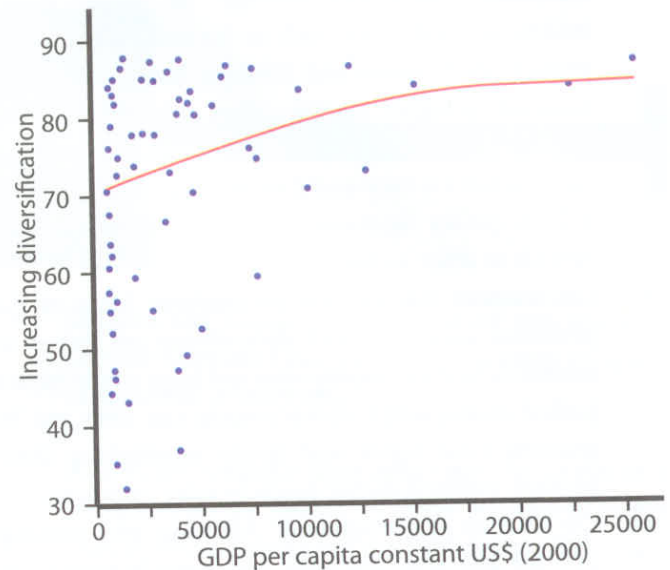
Figure 88.1 Correlation between increased diversification and income (2000-2004) (Source: World Development Indicators on line)

Strategies to diversify – up and/or out

Developing countries can basically diversify by broadening the product range (**up** the value-added chain) and/or the geographical scope of exports (**out** of a narrow range of exports and countries). This might be achieved is several ways:[1]

- **Gaining access to PTAs:** This would give preferential treatment to exports and allow for increased export revenue simply by selling more goods. The Lomé Convention between the EU and a group of African, Caribbean and Pacific countries (ACP states) allowed most primary goods from these countries free entry into the EU.[2]

---

1   See *Patterns of Export Diversification in Developing Countries*, World Bank International Trade Department, 2008 – passim

2   Look up the 'Banana war' for one of the most ridiculous trade disputes in the history of mankind.

DEVELOPMENT

613

- **Using existing resource for 'new' products**: A country endowed with certain natural resources can expand into non-traditional areas aimed at export markets. Roses in Ecuador, asparagus in Peru and fresh fruits in Uganda (which are basically airlifted to the EU!) are all fairly recent examples of this type of diversification.
- **Using historical ties**: Many previous colonies used historical ties to Europe to gain headway into export markets. Thousands of Indians in the UK meant demand for traditional foods – and these come from India.
- **Enhance the quality of goods**: This has been a reasonably successful strategy in some highly traditional 'cash crops' such as tea and coffee. Coffee growers in Tanzania and Ethiopia have increasingly focussed on higher-margin coffees rather than what is termed 'commodity coffees'. It has also been pointed out that there is very useful marketing value in getting a 'high quality reputation' in one sector as this then spreads to other sectors.[3]
- **Government intervention on markets**: Many of the problems associated with diversifying are linked to market failure, i.e. monopolies and large single buyers (called monopsony). Governments can help out by granting land rights and firmly establishing these property rights in owner deeds (= sales contracts).
- **Use FDI as a springboard**: Allowing and enticing FDI for low level assembly and manufacturing can ultimately allow workers the knowledge and expertise to start their own businesses later on. There are numerous success stories along these lines from India.
- **Expansion of the service sector**: While, again, largely contingent on resource endowments, developing countries can often attract tourism or provide various forms of 'outsourced' services such as call-back centres. Ruanda has incredible wildlife and India has millions of people who speak very good English. Work it out.

However …

It looks good on paper but in reality…well, think of it this way; if it were easy then it would have been done. The difficulties involved in diversification are numerous: lack of capital makes it difficult to move into secondary production; low education

and skill levels hamper moving into new product areas; poor and non-existent infrastructure make communications very difficult and costly; many of the poorest countries are land-locked or small island nations with high transport costs; complex bureaucracy coupled with political instability, corruption and crime stifle FDI; high tariffs and other forms of protectionism make exports more difficult; and finally, many of developing countries will be competing with each other for the same markets.

Summary & revision

1. **Diversification** of exports is a strategy for LDCs to develop by:

   a. **Reducing the impact of world prices** on export revenue

   b. **Adding to at-home value-added** and thus export revenue

   c. Increasing the **level of human capital** in the country

2. **Strategies** for diversification include:

   a. Gaining access to **PTAs**

   b. Evolving to '**new**' products

   c. Enhancing the **quality** of goods

   d. Increased **government intervention**

   e. Using **FDI** to move up/out

   f. Expanding the **services sector**

3. There are numerous **difficulties** in LDCs attempts to diversify:

   a. Lack of **infrastructure**/institutions, complex **bureaucracy** and **corruption** all serve to dissuade FDI

   b. Low **levels of human capital** make start-ups difficult

   c. **Lack of capital** and financing

   d. **Geographical hindrances** such as **climate** and being land-locked increases transport costs

   e. **Civil unrest** and **political instability** make it less attractive for FDI and tourism

---

3 Like the legend 'Made in Japan' which meant 'Crap' in the 1950s but by the 1970s was a stamp of high quality goods. In fact, my recent wristwatch, the Seiko Monster, does NOT say 'Made in Japan' and costs a good deal less than similar watches assembled wholly in Japan by Seiko.

4.5

89. Foreign Direct Investment (FDI) and Multinational Corporations

Key concepts:

- Definitions of FDI and multinational corporations (MNCs)
- Reasons for FDI in developing countries

One dollar of FDI is worth no more (and no less) than a dollar of any other kind of investment.
Dani Rodrik (Professor of International Political Economy, Harvard University)

Definitions of FDI and multinational corporations (MNCs)

Foreign direct investment is undertaken by *multinational companies/corporations* (MNCs)[1] - defined as companies that have production and/or services in several countries. FDI is defined as inflows which result in foreign ownership of fixed capital. This can come about in various ways:

- A company can build a *new plant* in another country or increase the size of existing operations – known as 'greenfield' investment.

---

1 Also known as multinational enterprises, multinational corporations, transnational corporations...and a few other variants. These terms are subject to fads basically, and seem to have the life expectancy of my marriages. Even the IB syllabus – after 11 years of using the bland 'multinational enterprise' – has returned to 'corporation'. We'll see what happens in the next syllabus.

- Mergers and acquisitions; a company can *join* with another company (merger) or *buy* a foreign company (acquisition or takeover).

- A firm (or financial institution) can *buy shares* in a foreign firm. To qualify as *direct* investment rather than portfolio investment the IMF defines FDI as any increase in foreign ownership of more than 10%.

The debate about foreign investment and MNCs, like that of globalisation, has diminished during the past few years, but there is still a bone of contention as to whether multi-national companies are predominantly pro- or anti-developmental. The focus here will be the *extent* of foreign direct investment, the *reasons* for it, and the *importance* of these capital flows for less developed countries.

Brief historical background

A number of factors combined during the 1980s and '90s to dramatically increase the flows of foreign capital to developing countries. The switch from inward oriented to outward oriented strategies invited more foreign capital. The GATT (WTO from 1995) rounds lowered tariffs and increased trade, which in turn encouraged multinational companies to seek lower cost opportunities for production, and decreased regulation on capital flows enabled them to set up subsidiaries abroad. The increase in free trade areas provided an incentive for international firms to 'tunnel under' tariff walls by establishing production and sales networks within the FTA barriers. Finally,

one should not underestimate the effects of the demise of the Soviet Union and the end of the Cold War in the early 1990s, as this cut off a great deal of politically motivated US and Soviet aid which in turn prompted more countries to look at foreign direct investment (FDI) as an alternative.

The amount of foreign owned capital in less developed countries doubled between the early 1970s and 1998, amounting to 22% of GDP. Figure 89.1 shows how foreign direct investment increased dramatically during the 1990s, and the dips during the 2001 and 2008 recessions.
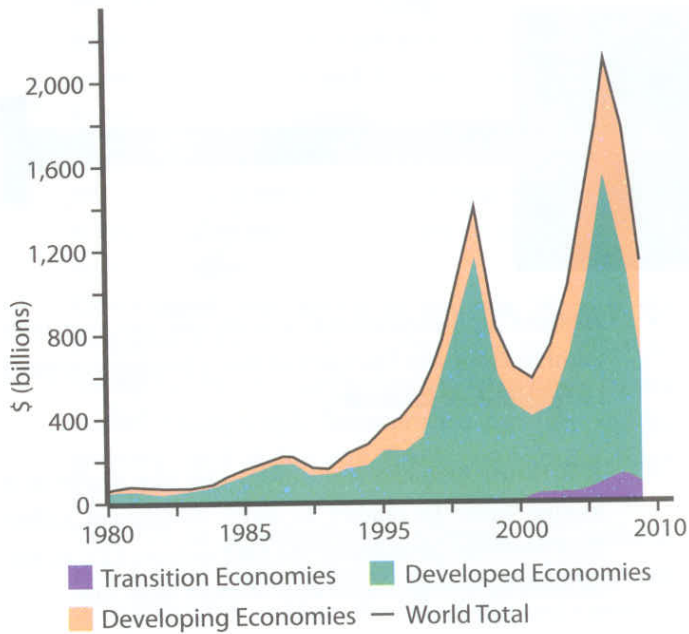
Figure 89.1 Total net flows of FDI 1980 – 2009
Source: World Investment Report 2010, page 2

By 2010 developing countries accounted for over 50% of all FDI inflows in the world. The top 5 receiving countries – the US, China, France, Hong Kong and the UK – accounted for over 50% of the inward flows of foreign direct investment. Encouragingly, about 45% of all 'green-field investment' is now flowing to developing nations.[2] Figure 89.2 shows the spread of total world FDI in 2010.

World	1,430,438
Low Income	13,017
Middle Income	501,236
Lower middle income	90,233
Upper middle income	411,003
Low & middle income	514,253
East Asia & Pacific	231,299
Europe & Central Asia	86,991
Latin America & Caribbean	117,368
Middle East & North Africa	25,688
South Asia	27,923
Sub-Saharan Africa	24,984
High Income	916,185
Euro Area	395,004

Figure 89.2 Total net flows of FDI 1980 – 2009 (Source: World Development Indicators 2012, page 368)

The marginalisation of Africa

While most other developing areas in the world saw a remarkable increase in FDI inflows during the latter part of the 1990s, Africa was the exception, receiving less than 1% of all FDI during 2000.[3] Africa – and primarily Sub-Saharan Africa – has been a notable outsider in the ongoing process of globalisation, receiving 1.7% of global FDI in 2009. One wonders why. The World Bank asked the same question in its report from 2002, *Globalization, growth and poverty* and considers three views:[4]

1. They have been unable to use their **comparative advantage** in abundant labour due to poor infrastructure, weak institutions, low education levels and political instability. This 'Join the Club' view[5] puts forward that, when these issues start to be resolved, Africa will join in the on-going process of global integration called globalisation.

2. Another view is that yes, Africa hasn't joined the club – and might well in fact have 'missed the boat'. Even if the institutional, infrastructural and political problems were largely solved, the fact is that globalised production patterns are becoming more set by the day and abundant African labour simply might not be needed.

3. The third view, mentioned in Chapter 79, takes a standpoint that many African countries suffer from a

3   *World investment report 2001*; UNCTAD, page XIV
4   *Globalization, growth and poverty*; World Bank 2002, page 38 ff
5   The report's words – not mine.

basic **geographic disadvantage**. Malaria and a number of other diseases limit tourism, as do harsh climates and lack of water in many instances. Huge distances and many landlocked countries with poor transportation networks and communications infrastructure have a resounding impact on trade and the costs of doing business.

Reasons for FDI in developing countries

There are a number of compelling reasons for firms to invest in developing countries, and it should be noted that such reasons deal with *revenue and profit* – not development. Firms are not in business to create welfare and equality but to make money for owners and shareholders. So while it is increasingly common for multinational firms to adopt a stance of 'corporate citizenship' and "corporate social responsibility", this is for the most part simply good marketing in the face of increasing opposition to multinational corporations. I am not being cynical or playing down any of the positive aspects arising from FDI and/or multinational companies, but simply pointing out that the aims of firms are not the same as those of development institutions.

Firms investing directly abroad in developing countries stand to gain on a number of counts:

- By setting up subsidiaries abroad, multi-national companies gain access to *lower cost factors* of production, primarily raw materials and labour.

- The 'global but local' argument: companies which sell internationally will not only have possible economies of scale but will also be able to better *monitor local consumer tastes and preferences* in order to adapt their wares internationally.

- Setting up both production and sales units within free trade areas will enable companies to *avoid tariffs* and other trade barriers.

- There are sometimes *logistical reasons*; a country well placed geographically can be used as a central supplier for surrounding regions.

- *Lower taxes* on labour, capital and profits will attract foreign direct investment. Also, several countries offering attractive profit tax rates – such as Panama and Belize – make attractive headquarters for MNCs, which can register in these 'tax havens' and pay far less profit tax.

- Firms producing environmentally hazardous goods might seek out countries with *low environmental standards*. (A textbook argument largely refuted by empirical evidence.)

Summary & revision

1. **Foreign direct investment** (FDI) arises when a company *builds a new subsidiary* plant or office (or adds on to existing physical capital) in another country. *Mergers, acquisitions* and the purchasing of a large proportion (over 10%) of the *shares* in a foreign firm also constitute FDI.

2. *Multinational companies/corporations* (**MNCs**) account for the vast majority of FDI – keeping in mind that both private individuals and governments can buy stakeholder shares in foreign companies.

3. **Reasons for direct investment** in LDCs include:

    a. *Lower factor costs* (primarily land and raw materials)

    b. Allowing MNCs to *scan local market tastes* and preferences – this is the 'global but local' argument

    c. MNCs can *avoid prohibitive tariffs* in LDCs

    d. *Logistical reasons* such producing in a region with large surrounding markets and/or proximity to good transport facilities

    e. *Low taxes* on labour and profits mean higher net profit margins for MNCs

    f. Firms might *avoid the costs of high environmental standards* (this argument does not stand up to scrutiny in reality)

DEVELOPMENT

617

90. Advantages and Disadvantages of FDI

Key concepts:

- The importance of FDI in developing countries
- Disadvantages of FDI
- A few words of warning
  - Comparing MNCs to GDP
  - The 'race to the bottom'

The importance of FDI in developing countries

The increase in FDI to developing countries during the beginning of the 1990s corresponded with a decline in official development assistance and other forms of aid. Thus, while total capital flows to less developed countries grew throughout the 1990s, the *composition* of these flows changed dramatically, as Figure 90.1 shows.
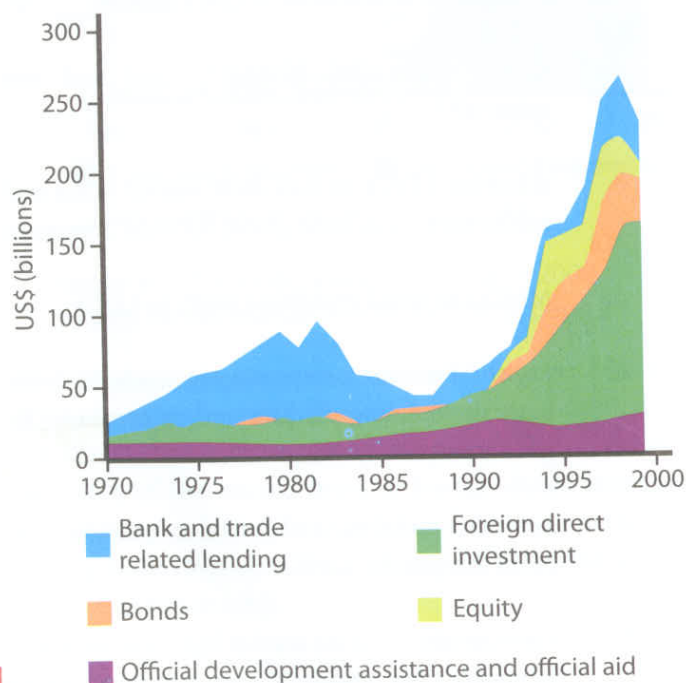
Figure 90.1 Total net flows of capital to developing countries 1970 – 1998

Source: Globalization, growth and poverty; World Bank 2002, page 42

Official flows decreased by more than 50% and private capital flows are now the major source of inflows to developing countries. These flows of foreign capital can have a considerable impact on developing countries. While the results of FDI are still debated, the following points illustrate the *potential* benefits of foreign direct investment in less developed countries:

Filling a savings-investment gap

Perhaps the main benefit of foreign direct investment in a developing country is that the gap between domestic savings and domestic investment can be filled. Low income countries will have low savings ratios (as a percentage of incomes) since there is basically very little to spare for households living on the bare minimum. This makes available funds for investment very scarce and impedes investment. How to fill the gap between savings and investment?

Before you read on, go back and revise the circular flow of income in Chapter 36. Notice that the sum of injections equals the sum of leakages, i.e. $T + S + M = G + I + X$. Now go back and revise the balance of payments in Chapter 70, paying particular attention to how a country might have a current account deficit, i.e. $X < M$. Recall that a current account deficit is not 'bad' depending on where the inflows in capital account are coming from.

Assume, for simplicity's sake, that an economy has a perfectly balanced budget, in other words that government spending equals tax revenues; $G = T$. Now, what if savings are less than total investment in the economy? Say that total savings are $US500 million and total investment is $US600 million:

$$T + 500 + M = G + 600 + X$$

Since the government budget is perfectly balanced, we can strike T and G on each side:

$$500 + M = 600 + X$$

Now, skipping all the stuff you will pester your maths teacher with, we can see that *imports* are going to have to *exceed exports* by \$US100 million. And this should be enough for you, but I'll slam in the final nail: X – M = net exports, which means that there is a **current account deficit** of \$US10 million. Consequently, there is a net inflow of \$US100 million on **capital/financial** account, which explains how the country has just been able to finance \$US100 million over and above domestic savings. While a proportion of these inflows are likely to be loans, this simple example helps to illustrate the importance of developing countries being able to attract foreign investment.

Export earnings and foreign currencies

FDI provides an increase in foreign currency in the host country when the investment takes place which contributes to **export earnings** in foreign currencies. Naturally a lot of FDI is a 'one off' and there will be successive outflows from host countries as profits are repatriated back to the country where the MNC is headquartered. Figure 90.2 shows the strong correlation between rising FDI flows to LDCs and export earnings during the boom period in the 1990s.
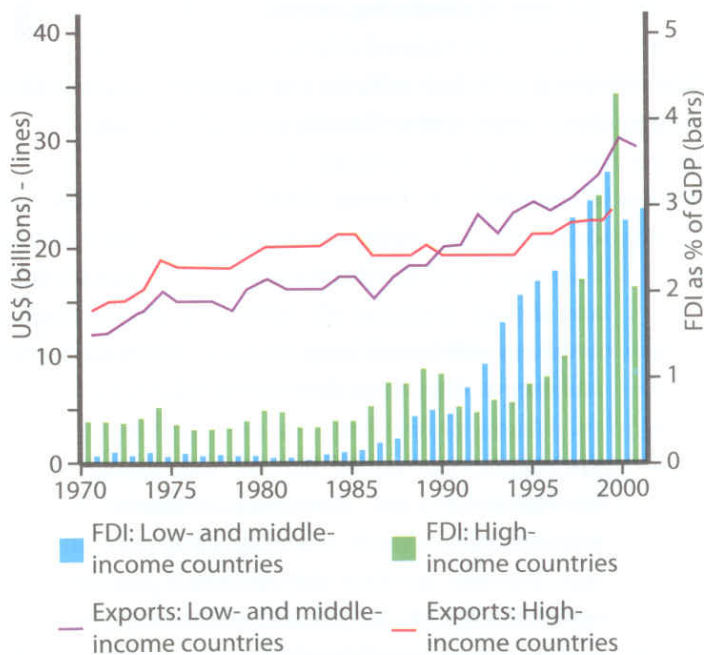
Figure 90.2 Increase in FDI and exports in LDCs, 1970 – 2002
(WDR 2005, page 65)

Employment and tax revenues

Foreign firms create jobs and **provide tax revenue** for host governments in less developed countries. Labour, income and profit taxes will all contribute to much-needed spending on infrastructure, health care and education. When US and European auto makers increased FDI to Mexico in the 1980s, employment in this sector reached over 350,000 jobs – second only to the petroleum industry.[1]

D-side effects on the economy

Increased economic activity will have **secondary effects** on the host economy, as local parts manufacturers, builders, service providers etc. are needed. Additionally, the increase in incomes for workers will fuel demand for other goods and help generate an increase in aggregate demand. (HL should recognise these *multiplicative effects*.) If the host government in a developing country has offered various forms of tax benefits, one needs to take into consideration whether the tax losses are offset by the wider gains of increased jobs, knowledge and technology transfers and the future gains of being regarded as a good place to invest. This is not an easy task to assess as pointed out on several times by the UNDP.[2]

Increased competition and productivity

The presence of highly efficient multinationals can have a stimulatory effect on **competition** and therefore efficiency in developing countries, ultimately pushing down costs for domestic firms and improving resource allocation. There are also scale benefits to be considered. Several studies have shown that FDI does seem to have a positive effect on productivity – the 'if' being that the effect of FDI on productivity is strongly linked to the level of human capital in the recipient country; FDI in countries with higher levels of human capital yield higher productivity increases.[3]

Transfers of knowledge and technology

Foreign direct investment to a developing country is often not a simple transfer of money, equipment and managers, but a **transfer of technology, knowledge and skills**. By increasing the level of sophistication in the use of technology, a developing

1 Peterson Institute for International Economics; *How does FDI affect host country development?*, page 283
2 See, for example, the Human Development Report 2005, page 169)
3 See, for example, *Foreign direct investment in developing countries: Leveraging the role of multinationals*, 2002, Norwegian Research Council, University of Oslo, page 6 ff

DEVELOPMENT

country will ultimately increase its technological level through training programs and 'learning by doing'. Skilled labour from abroad can educate local workers and help ease a skill-constraint found in many less developed countries. A report from 2008 by the World Bank broadly corroborates the link between FDI and spread of technology – see figure 90.3.
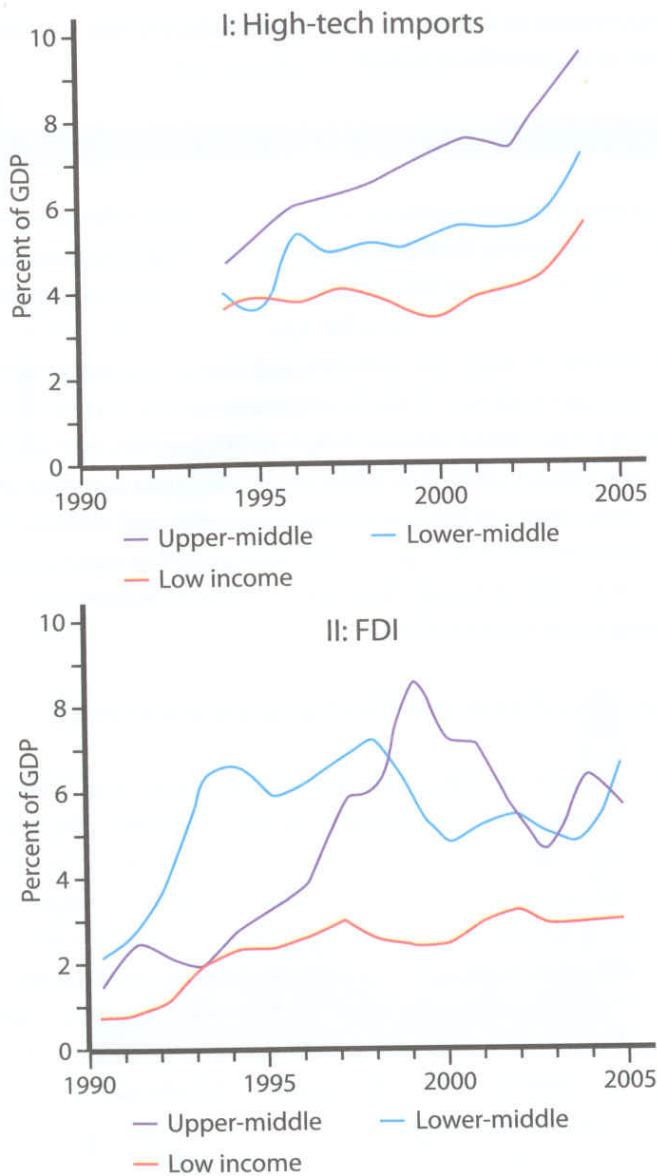
Figure 90.3 FDI and increasing technology spread in LDCs

Consensus?

Yes, of course there is disagreement and a lot of contradictory data – many of the possibilities above will, of course, not present themselves to *all* recipient countries and there is considerable scope for criticism. Yet the World Bank has found that, while there are indeed a number of cautionary notes within the field of FDI, there is increasingly strong evidence of the *positive growth effects* of FDI inflows. Point in fact; FDI in developing

countries would appear to have a much stronger effect on economic growth than the overall level of investment. In other words, **foreign** investment seems to have a greater effect on growth than **domestic investment**. Instead, other domestic factors such as education and infrastructure seem to have a greater impact on growth than domestic investment.[4]

Disadvantages of FDI

The controversies over multinationals and FDI arise on two basic levels; the ideological and the functional.

The **ideological split** arises over the question of whether free market private enterprises are good or bad for development in general.

- **Proponents** of foreign direct investment and multi-national companies are primarily of the free-trade/liberal school.

- **Opponents** of multinational companies look at the power relationship between developing countries and view multi-national companies as a threat to the cultural/political/social/economic identity and strength of developing nations.

The **functional split** deals with the pro-developmental aspect of foreign direct investment and multi-national companies.

- **Proponents** of foreign direct investment view the transferral of capital in accordance with classical economics; capital will seek out the highest rate of return and allocate resources efficiently, whereby developing countries will benefit from increased investment, jobs and output. This is pro-developmental.

- The **opponents** would counter that foreign direct investment in less developed countries in fact empowers the already powerful companies and the rich countries in which they are based. Multinational companies benefit the few and widen inequality gaps, and are able to exploit low labour cost countries while diverting profits homeward rather than reinvesting them in the host country.

While FDI and multinational companies clearly can benefit developing countries, there are a number of points of criticism which most definitely deserve notice:

4 *Globalization, growth and poverty*; World Bank 2002, page 43

- Multinationals are not prone to reinvest profits. Thus the initial investment will be countered in the long run by outflows of profit, and *adversely effect the current account* in balance of payments.

- It is quite possible that a multinational presence *forces local firms out of business* as these smaller firms will not be able to compete with the larger rival. This could in fact lower total saving in the economy due to the loss of jobs and income.

- It can be argued that *fixed capital* use by multinational companies is inappropriate for developing countries as such production will not be labour intensive enough to have an impact on unemployment. Furthermore, capital will create future demand for imports such as spare parts and specialised labour for maintenance and renovation.

- Multinationals have also been accused of using monopoly power to create demand for *inappropriate products* aimed – via heavy marketing – at a relatively small clique of well-to-do people.

- Multinationals are of such *economic power* that they in effect have considerable leverage in getting good terms from developing countries which desperately need the capital inflows and employment opportunities provided by multi-national companies which set up subsidiaries. If tax concessions are granted, there will be less of an improvement on the developing country's tax base and tax receipts.

- An ever louder argument – enforced by increasing evidence that it in fact is quite a common practice – is that firms can quite simply avoid a lot of profit taxes by a method known as *transfer pricing*. Say a German manufacturer has a subsidiary company in low-tax Panama producing components for German refrigerators to be sold on the EU market. The Panamanian subsidiary manufactures the components for USD5 apiece but when 1 million components are shipped to Germany the Panamanian subsidiary bills the German headquarters (HQ) for USD10 apiece. In the Panamanian subsidiary's profit statement, *revenue* has doubled and in the German HQ profit statement *costs* have doubled! A tax saving double whammy; taxable profits (which are basically total revenue minus total costs) have increased in low-tax Panama and decreased in high-tax Germany. The end effect is that the money stays in the MNC but total taxes payable have decreased. You probably guessed that countries that are losing taxes have legislators sharpening their knives to limit the practice.

The debate on FDI is far from over. The next decade will surely show results that both prove and disprove both sides. Hopefully the heat and divergence of commentators' opinions will diminish, allowing for a more level-headed analysis of the outcome. I bring my own straw to the stack, by scaling away two of the most common but largely irrelevant – and erroneous – points put forward within the issue of multi-national companies' power.

A few words of warning

There are a few frequently reoccurring 'truths' that have rattled my cage, yanked my chain and scuffed my rhubarb patch for some time now. The two subheadings below can well be avoided as they will add nothing to your ability to deal with the IB exam. Basically I am just letting off some steam. It's cheaper than therapy.

Comparing MNCs to GDP

You might want to revise national income accounting before reading this. Specifically, you will need to have a firm grasp of the output method of assessing GDP and the **importance of avoiding double counting**. (See Chapter 37.) When you have done so, you will be able to grasp a fundamental economic fallacy that is put forward all too often.

You see, the inability or rather lack of propensity of a number of economists, NGOs and anti-globalists to understand the error in double-counting has led to a most common and seriously flawed comparison, where *total revenues* of large multinational firms are *compared to the GDP* of developing nations. This comparison is something like comparing apples and…wildebeest, yet amazingly enough of it has been done by a number of very well-educated and knowledgeable people. Quite frankly, *too* well-educated not to know that what they are doing is erroneous, which means that they are being deceitful rather than ignorant – a far scarier offence.

These apparently well-meaning but rather woofly[5] commentators often state that multinational companies are becoming more powerful than governments. This is either heavily inferred or explicitly stated. This outrageous off-hand use of economics usually includes a list of countries and multinational companies, such as that in Figure 90.3. The accompanying text would then read something like "The size of many multinationals now exceeds the total income of many countries…", and continue with a line of reasoning that Exxon Mobile, with total sales of $US275 billion in 2011, has more economic power than Angola, Ghana, Ukraine and Sierra Leone put together – which had a combined national output of $US311 billion in 2011.

What an amazing fact! What a stunning and eye-opening comparison! What utter **nonsense**.

Figure 90.4 Largest companies by sales compared to GDPs of selected countries 2011

Company	Sales (= revenue) (billion $US)	Profit (billion $US)	Country	GDP (billion $US at current values)
Wal-Mart	408	14	Angola	100
Royal Dutch Shell	278	12	Ghana	37
Exxon Mobile	275	19	Ukraine	165
British Petroleum	239	16	Greece	303
Toyota Motor	210	-4.5	Sierra Leone	9

Sources: Forbes Fortune 500, at http://www.forbes.com/lists/2010/18/global-2000-10_The-Global-2000_Sales.html and IMF World Economic Outlook (WEO) Database, 2012.)

Let's have a closer look at the issue without giving it away. (Just keep thinking to yourself why it would be absolutely idiotic to compare the *GDP* of a nation with the *profit* of a firm.) Assume a small country which has three large firms that produce all output in the economy; Megafirm, Superfirm and Maxifirm.

5 You will have noticed my use of 'woofle', or 'woofly' by now. You will not find woofle in the dictionary as I have made it up. It is a composite term like stagflation (**stagnant** + **inflation**). **Woofle** is comprised of *woolly* (as in confused or muddleheaded) and *waffle* (hazy, imprecise and evasive). I figure that if enough of us economists use the term, we can get it into the Oxford Dictionary. However, since you heard it here first; Woofle'.

Say that Megafirm sells $US100 worth of goods to Superfirm (illustrated in Figure 90.4 and Figure 90.5), which Superfirm then uses as intermediate goods in the production of $US400 worth of goods sold to Maxifirm in round 2. Stop and calculate total revenue and total output; when Superfirm sells goods to Maxifirm, revenue is $US400, but total output is only $US300 as the $US100 dollars worth of goods used as factor inputs by Superfirm **must be deducted to avoid double counting!** The same holds true as these intermediate goods – factor inputs – flow around the production system during rounds 3 to 5; total revenue will of course be higher than total output value during each round.

Figure 90.5 Revenue and value-added in an economy ($)

Rounds of selling and buying in a simple economy	Total revenue	Total money value of output (value-added) at each stage = contribution to GDP
Round 1: Megafirm sells goods to Superfirm for $100	$100	$100
Round 2: Superfirm sells goods to Maxifirm for $400	$400	$300 ($400 – $100)
Round 3: Maxifirm sells goods to Megafirm for $1,000	$1,000	$600 ($1,000 – $400)
Round 4: Megafirm sells goods to Superfirm for $2,000	$2,000	$1,000 ($2,000 – $1,000)
Round 5: Superfirm sells goods to Maxifirm for $4,000	$4,000	$2,000 ($4,000 – $2,000)
Round 6: Maxifirm sells goods to final consumer for $8,000	$8,000	$4,000 ($8,000 – $4,000)
Sum totals:	$15,500	$8,000 (= GDP)
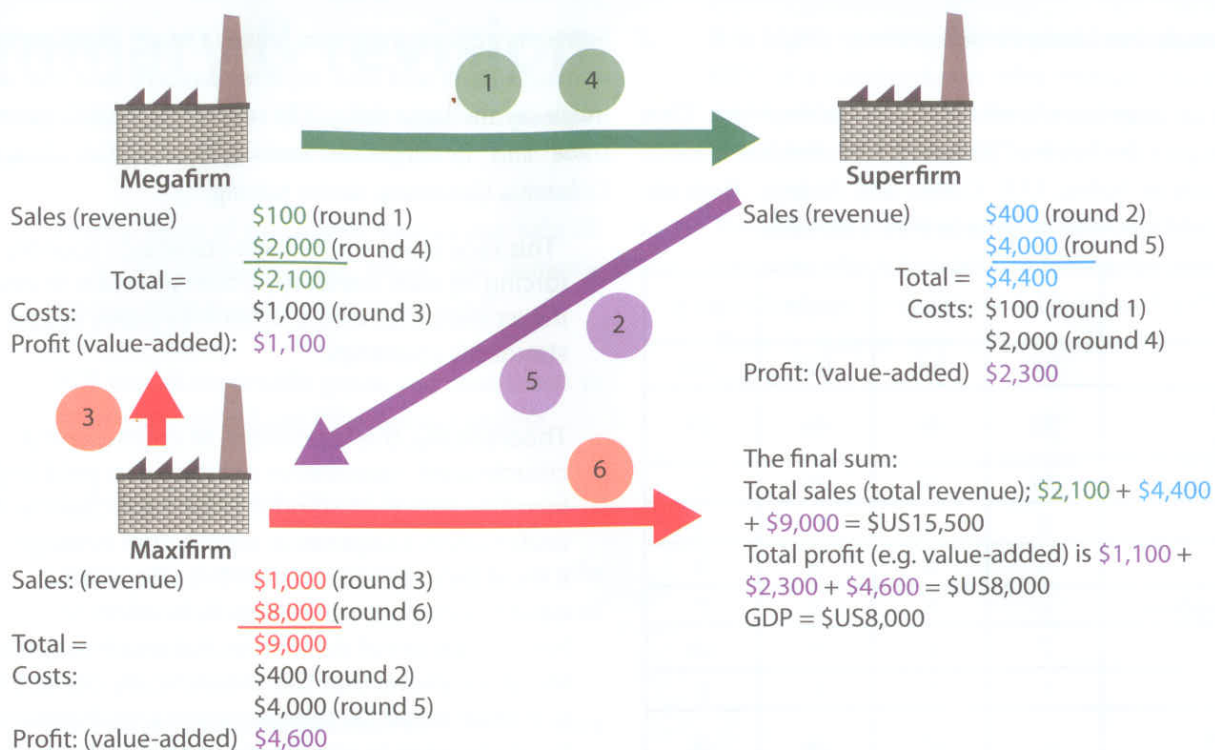
Figure 90.6 A simple 3-firm economy; revenues, profits and GDP (value of final output)

This is exactly what happens in real economies; a large portion of output is used as inputs in other firms. When Maxifirm ultimately produces an **$US8,000 Widget** which is sold to an end-user in round 6, the final sale is the same as the sum of value-added, i.e. a **GDP of $US8,000**. Now here's the interesting part: if you look at the profit and loss statement of each firm, you will discover that summing up *each firms'* **profit** *gives exactly the same value as GDP!* Total revenue is entirely dissimilar to de facto output value. Bluntly put, one simply *cannot* compare total sales (revenue) of firms with the total value of output. Total revenue includes all the costs of inputs and value added – while GDP is only the (rough) equivalent of total value added.

Now go back and look at Figure 90.3 and spot the flawed reasoning in comparing firms' total sales with countries' national outputs. The appropriate – but still rather meaningless – comparison would be **between companies' profits and GDP**.[6] It then turns out that ExxonMobil in fact is considerably smaller in 'economic power' than Ghana. Toyota Motor Company ran a USD4.5 billion loss in 2011. In other words, the 'Toyota economy' was in fact 'weaker' than Sierra Leone – one of the poorest countries in the world.

---

6   Note that in reality there will not be a **perfect** match between corporate profits, value added and GDP. Firms will have additional costs and income other than those associated purely with the sales of goods. Profits and GDP are the **closest** comparisons one can make in a very silly exercise which compares *companies* to *countries*.

The 'race to the bottom'

There is a common argument – or anti-globalisation rant as it frequently turns in to – that developing countries have an incentive to lower standards on environment, pollution, and labour in order to attract FDI. The logic is that regulations on these standards, while highly socially desirable and beneficial, raise costs for foreign firms and thus dissuade investment flows. Proponents of this opinion put forward that MNCs will scour the developing world in search of countries with the weakest/lowest (or unenforced) rules and regulations on environment, labour and taxes in order to incur the lowest possible costs in production. LDCs thus become 'pollution and labour exploitation havens' whereupon rich countries can in effect export industries with high negative externalities.

There is very little hard-line evidence *for* this argument and a great deal of empirical evidence *against* it. One can sum up the empirical counter-argument in one word: Africa.

The argument underlying a 'race to the bottom' presupposes a number of things in the LDC:

1.   A weak or corrupt government that is willing to forego the interests of its people in order to attract FDI.

2.   Weak institutions unable to uphold labour, environmental and tax legislation.

3.   Extremely low labour costs.

Historically, all three ingredients are to be found in Africa. Thus we would expect the brunt of global FDI be directed towards countries such as Sudan, D.R. Congo and Nigeria. Here are what 'greenfield' investment flows to Africa look like:

Host region/economy	2007	2008	2009	2010
World	100	100	100	100
Developed economies	52	46	46	49
European Union	39	34	30	31
France	5	4	3	3
Germany	4	4	3	3
United Kingdom	6	5	8	7
United States	7	6	9	10
Japan	1	1	1	1
Developing economies	42	47	48	45
Africa	3	5	5	5
South Africa	-	1	1	1
Latin America and the Caribbean	7	7	9	8
Brazil	1	2	2	2
Mexico	2	2	2	2
Asia	32	35	34	32
West Asia	5	7	7	7
South, East and South-East Asia	27	28	27	26
China	10	9	8	8
Hong Kong, China	1	1	2	1
India	6	6	5	6
South-East Europe and the CIS	6	7	6	6
Russian Federation	3	4	3	3
Total number of cases	12,210	16,147	13,727	4,104

Figure 90.6 Global investment flows for the world; 2007 to 2010 (Source: World Investment Report 2010, Chapter I, page 4)

Somebody is losing the race it would seem. Skipping hundreds of World Bank and IMF reports of recent date that all pretty much say the same thing, I let one of the world's most famous trade and development economists, Jagdish Bhagwati of Columbia University, do the talking:

"This race refers to the high-standards country forcing its own standards down to retain or attract investment that would otherwise move to low-standards countries.

Theoretically, this is possible, of course, and a coordinated, cooperative solution can yield higher incomes and even standards for each country than under a non-cooperative equilibrium (though, it must be emphasized that this cooperative equilibrium does not generally involve harmonization of standards). But much empirical literature shows that this theoretically-correct fear is of little empirical relevance since multinationals typically seem to use the more environmentally-friendly technology even when not required, and there is little evidence either that governments actually lower standards in order to attract investments."

Source: This is the text of the Keynote Address delivered at the University of St. Gallen, Switzerland, on 25th May 1998, on the occasion of the International Management Symposium at which the 1998 Freedom Prize of the Max Schmidheiny Foundation was awarded.

Summary & revision

1. Foreign direct investment (**FDI**) has numerous possible **benefits** for developing countries:

    a. Inward flows of investment monies can help fill a '*savings-investment gap*' where low domestic savings stifle available investment funds

    b. FDI creates exportable goods and contributes to aggregate demand and thus *national income*

    c. FDI creates *employment* and *tax revenues*

    d. There are potential *secondary effects* on the economy such as work for sub-contractors and service providers

    e. HL: the investment inflows can have *multiplicative effects*

    f. FDI has been shown to increase both *competition* and *efficiency*

    g. There are benefits in the *transfer of knowledge* and *technology* via FDI

2. There are possible **negative effects of FDI**:

    a. FDI can be a 'one-off' aggregate demand stimulus – *profits are frequently repatriated* rather than re-invested in the host country

    b. Large foreign firms can *out-compete domestic firms* causing bankruptcies and loss of jobs

    c. FDI can be *inappropriate for LDCs* if there is a high level of fixed capital and resulting low demand for labour but high demand for foreign parts and expertise

    d. MNCs can be powerful enough to *force host nations into concessionary behaviour* such as granting tax relief

    e. MNCs can indulge in *transfer pricing* – a practice that lowers the overall taxes paid by multinationals by way of accounting for more profits in low-tax countries than in high-tax countries

3. It is highly questionable whether one can compare the GDP of a country to the sales *revenue* of firms. A more appropriate – but still rather meaningless – comparison would be to compare GDP and *profits*.

4. There is little evidence that MNCs have located to countries which 'compete' in having the lowest possible standards on environmental, labour and tax regulations. There is simply very little evidence for a 'race to the bottom'.

DEVELOPMENT

4.6

91. Classification of Foreign Aid

UNESCO

United Nations
Educational, Scientific and
Cultural Organization

Key concepts:

- Reasons for aid
- Types of aid
  - o Bilateral and multilateral aid
  - o Donor or grand aid, soft loans
  - o Humanitarian aid
  - o Official development assistance (ODA)
- ODA data
  - o Non-government organisations (NGOs)
  - o Tied aid
- Aid data for two LDCs

"Charity is no substitute for justice withheld." Saint Augustine

Reasons for aid

Quite frankly, the basic answers should be "because they need help". Short term – immediate – aid is often in the form of emergency aid relief due to natural disasters, civil war or political unrest which put people in LDCs that are already living on the margins in real harms way. Most types of aid aim at medium to long term development such as *transferring expertise and knowledge* in the form of foreign experts and technology; increasing *investment opportunities* (see 'savings-investment gap' in Chapter 90); providing *development basics* such as roads and telecoms infrastructure; and generally helping to create the *institutions* necessary for development to take place.

Types of aid

Foreign aid consists of outright grants (donor aid), soft loans and tied aid. This is mostly undertaken by official aid agencies, multilateral aid agencies and non-government organisations (see below). Also, while some literature includes debt relief – i.e. writing off debilitating loans for heavily indebted poor countries – I exclude this from the discussion below and give it a heading of its own further on.

Bilateral and multilateral aid

Bilateral is literally 'two sides', so in the case of aid it is one country giving support to another country directly. **Bilateral aid** is commonly done via direct money support, shipments of machines and tools, emergency aid such as food, medicines and temporary housing, and experts who are sent to help in building dams or setting up road projects. Bilateral aid is often targeted at countries due to political considerations and/or historical ties – see below.

Conversely, multilateral means 'many sides'. When the World Health Organisation, UNESCO[1] (both UN agencies) and the World Bank disburse funds which come from donor countries one speaks of **multilateral aid**. A multilateral organisation is thus allocating funds rather than the country providing the funds.

---

1   United Nations Educational, Scientific and Cultural Organisation

DEVELOPMENT

Donor or grant aid, soft loans

There is a somewhat bewildering variety of terms for aid – basically it's another umbrella concept (e.g. 'covers a lot of different concepts'). Aid can be defined as any non-commercial and non-military flow of funds to a developing country. **Grant aid** is a 'gift' of money, goods, capital etc, for which there is **no reciprocal agreement** or strings attached. When money is lent on favourable terms – i.e. at interest lower than market rates and/or with longer amortisation periods – then one speaks of **soft loans**. Such loans are frequently managed by the International Bank for Reconstruction and Development (IBRD) – more commonly known as the World Bank – and other national development banks.

Humanitarian aid

> "It is no good to lecture the dying that they should have done better with their lot in life." Jeffrey Sachs

This type of aid is primarily short run and 'immediate' – actions in humanitarian aid aim to help those in extreme situations of suffering during and after crises such as civil wars or floods. This is perhaps the least politicised of aid types in that it is based on the general principles of humanity and political/cultural neutrality. Aid agencies commonly look at three points of 'emergency response':

- Material relief such as water, tents and basic medicines
- Emergency food aid distribution
- Coordination of relief and logistics

Official development assistance (ODA)

Official aid, often referred to as official development assistance (ODA) is the sum of bilateral and multilateral aid, thus comprising all aid forms not coming from charities such as Oxfam and religious organisations. ODA is thus administered by governments or government agencies and goes directly to other countries or via multilateral organisation such as the IMF and UN. Figure 91.1 shows total ODA over a 20 year period – the most notable trend being the decrease of aid disbursement in the wake of the financial crises in 2010.
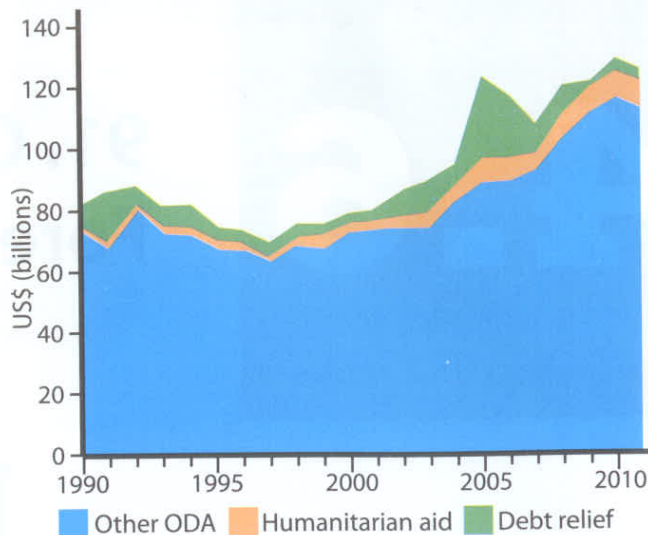
Figure 91.1 Official Development Assistance – totals for period 1990 to 2011 (Source: OECD online database at www.stats.oecd.org)

ODA data

The United Nations (UN) has a target aid objective that wealthy donor countries should donate a minimum of 0.7% of GNI per year. Figure 91.2 shows a selection of the wealthiest countries in the world – few of which meet the UN aid target. In fact, only 5 countries have fulfilled the UN objective; Luxembourg, the Netherlands, Norway, Sweden and Denmark. Another notable item is that while US aid is the highest aid contribution (USD in absolute money terms), it is the proportionately fourth from the bottom. Germany as number two in absolute terms is second from last in relative terms.

A pallet US Aid is delivered
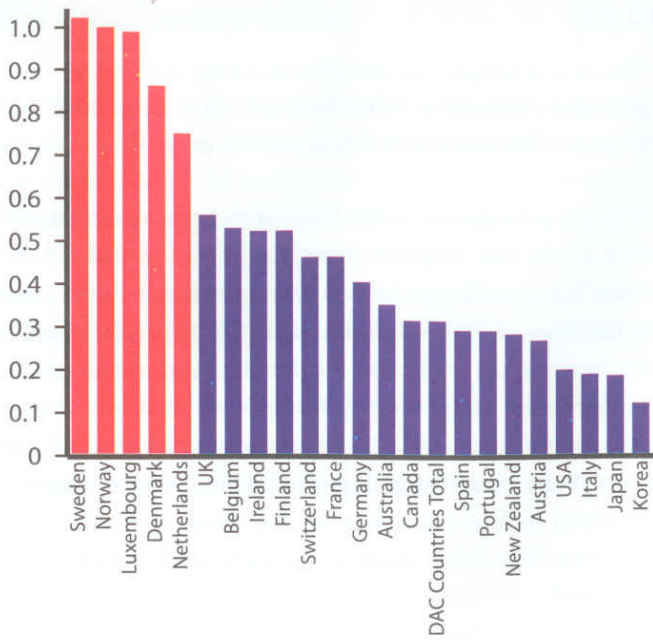
Figure 91.2 Official Development Assistance as a percentage of GNI, 2011 (Source: OECD statistics at http://www.oecd.org/dataoecd/42/61/31504039.pdf)

The richest donors (the 23 DAC donor countries[2]) disbursed USD129.4 billion in aid during 2011. The EU and the 27 member countries is the world's largest aid donor at USD66 billion – more than half of all ODA during 2011. This doesn't mean a lot until one puts it into perspective. Let us look at this USD129 billion in comparative terms, e.g. as a percentage of, say, government spending on farmers and weapons.

- Total *OECD subsidies* and ODA: USD366 billion (2010) in support for OECD farmers which makes up about 18% of total farm income in the OECD.[3]

- World *spending on arms* and ODA: total military expenditure in the world in 2011 was an estimated USD1,600 billion.[4]

I can't remember where I read it, but there's a great and poignant (= moving) quote on the 'guns or butter' issue in developing countries; "There never seem to be enough trucks to transport food to famine stricken areas – but somehow there are always trucks available to transport arms." I don't need to spend 45 minutes going through data from UNHCR to know the sad truth of this.

2   This is the Development Assistance Committee of the OECD – 23 OECD countries plus the European Commission.
3   OECD, *Agricultural Policy Monitoring and Evaluation 2011*, page 17 and World Development Indicators 2012, page 33.
4   Stockholm International Peace Research Institute at http://www.sipri.org/research/armaments/milex/resultoutput/world

Can you say "opportunity costs"?

ODA; Official Delivery of Arms

Non-government organisations (NGOs)

A non-government organisation, or NGO, is a voluntary organisation which is funded by contributions rather than government funds. These organisations span not only the globe but a wide range of more or less special interests. Examples of some of the larger include Oxfam, the World Wildlife Fund for nature (WWF), Christian Aid, Greenpeace, the Red Cross/Crescent, and around 38,000 others. Some 30,000 of these NGOs were created during the 1990s.[5] The phenomenal increase in NGOs during the 1990s of course coincides with the ability of a small number of people with special interests to reach an enormous number of people around the world via the Internet.

The proliferation of NGOs has had considerable positive impact on development issues and policies over the past decades, and their influence is growing. A number of benefits can be seen for development in this:

- Many NGOs conduct and **publish studies** which bring to light a number of issues which might otherwise be sidelined. Oxfam reports are increasingly used as a reference on the effects of trade barriers in developing countries.

- The ability of NGOs to coordinate efforts across country borders makes them a force to be reckoned with as **pressure groups** towards government policies and the activities of multinational companies.

- Many of the development NGOs operate in developing countries; setting up irrigation projects, operating field hospitals, delivering emergency aid relief and so forth. The fact that most of these NGOs are **politically**

5   *Legrain*, page 202

DEVELOPMENT

unaffiliated means that local people are more prone to trust them, which gives them wider margins to operate in and increases the efficiency of development projects.

- The vast majority of people working for NGOs are volunteers. Many are from developed countries and will share their experiences at home, helping to spread hard-earned and first-hand knowledge of development issues.

However...

There are a few negative aspects of NGOs. I have several friends who are long-standing members of Greenpeace and other environmental NGOs while I also have friends who are card-carrying members of conservative democratic parties. Guess which of them get invitations to elect leadership and/or to speak out at conventions to set policy guidelines? Correct; NGOs are often run by a non-elected clique or core of dedicated leaders surrounded by a much larger group of supporters and contributors. Some sharp criticism of the lack of democracy and accountability in many NGOs has been put forward – notably by officials at the receiving end of NGO hostility. The former Director-General of the World Trade Organisation, Mike Moore, and Philippe Legrain (also formerly WTO), pose a number of barbed questions:[6]

- Why don't NGOs go the traditional democratic route of ballot boxes and elections? Is this due to the realisation that broad support is in fact simply lacking?

- Why should special-interest groups consisting of unelected people and representing a very small minority be granted a seat at international conferences together with duly elected democratic representatives? Is it reasonable that 2.5 million members of Greenpeace can make demands on international issues with the same weight as the Indian parliament which is elected by 600 million voters?

- Why do many NGOs seem to be exempt from transparency demands such as open financial statements and meeting protocols? Especially when many of these NGOs themselves have criticised official and multilateral development organisations and multinational companies of operating behind closed doors.

---

6 *Moore*, page 187 ff and *Legrain* 202 ff

Tied aid

"There is a simple method for tackling the waste of money associated with tied aid: stop it in 2006." (Human Development Report 2005, page 9)

Bilateral aid is frequently in the form of reciprocal agreements between donor and recipient countries, known as tied aid. It's a bit like "I'll scratch your back if you scratch mine", but more along the lines of "We would be happy to help your country build a new airport – oh, and incidentally, you might also be in the market for some very nice fighter jets that my country happens to produce!" OK, I'm being facetious again, but this is in fact pretty much what happened when Sweden offered a few hundred million US dollars in assistance to South Africa in the 1990s. You didn't think SAAB only made ugly but safe cars did you?

Tied aid is in fact seldom as nasty as I've portrayed it above, but I can resist anything except temptation.[7] Most tied aid is in fact simply a form of agreement between donor and recipient nations, where loans or direct money to the recipient is to be used on goods and services from the donor country. Since the donor country sees an increase in exports and the aid is disbursed, one might say that something of a win-win outcome is possible.

However ...

It depends, for the most part, on what kinds of strings are attached and where the aid is going – as in all cases of aid. One major criticism of tied aid is that the cost of the reciprocal goods and services purchased by the developing country can be far higher than when purchased on a competitive market. The OECD estimates that the average cost of such tied-aid goods is 15 – 30% higher.[8] The UNDP estimated that yearly losses due to tied aid amount to between USD5 to USD7 billion – enough to fund primary education in all developing countries.[9] Tied aid constituted 14% of all ODA in 2009 with the US (72% tied), Greece (77%) and Italy (92%) topping the list.[10]

---

7 ...and misquotations. This was of course said by Oscar Wilde.
8 OECD report; *Reaching our development goals* 2008, page 16
9 UDR 2005, page 119
10 See OECD online at *Untying aid* and *Where does the money go?*, Brookings Global Economy and Development, Working Paper 21, June 2008, William Easterly and Tobias Pfutze, page 17

Aid data for two LDCs

Aid monies can have a considerable impact on a developing nation. One of the world's poorest nations, Liberia, receives 176.8% of GNI in the form of aid while comparative figures for Namibia and Uruguay are 2.1% and 0.1% respectively.

Rather than a tiresome string of data and tables, I urge you to go to the excellent OECD website and have a look at aid recipients and donors:

http://www.oecd.org/countrylist/0,3349,
en_2649_34447_25602317_1_1_1_1,00.html

Summary & revision

1. **Bilateral** aid is aid given directly from one donor country to a recipient country. **Multilateral** aid is channelled through development organisations such as the World Bank and UNDP.

2. **Donor** or **grant aid** is a 'gift' of money, goods, capital or expertise.

3. **Soft loans** can be considered a type of aid, in that loans are given with longer repayment times and/or lower than market interest rates.

4. **Humanitarian aid** is immediate action taken to alleviate human suffering as the result of natural disasters, war and civil unrest.

5. Official development assistance (**ODA**) is the total aid administered by governments – whether bilateral or multilateral.

6. Non-government organisations (**NGOs**) are voluntary non-tax-funded organisations with relatively specified areas of focus. Oxfam, WWF and Greenpeace are NGOs.

7. **Tied aid** means that the receiving country is contractually bound to some form of reciprocity in receiving aid – often in the form of purchasing goods from the donor country.

DEVELOPMENT

92. Evaluation of Foreign Aid

Key concepts:

- The aid debate
- Aid and development – a very tricky evaluation
- Aid vs. trade

"Only a life lived for others is a life worthwhile." Albert Einstein

The aid debate

Few observers – economists or otherwise – would be willing to condemn aid across the board (see 'A little depth: Views of development economists' further on). Yet a number of questions persist as to the efficiency and desirability of aid:

1. Is aid going to those most in need or is aid given primarily for other considerations, e.g. for **political**, **military** and **self-interest** reasons? It is evident that donor countries allocate aid primarily to countries with which there are previous – often colonial – ties, and also where there are strong strategic arguments supporting aid.[1] For example, the world's largest donor nation, the US, has focused aid on Southeast Asia to create a defence against communism in the 1960s, and later shifts in aid were directly caused by changes in American considerations to strategic and political policies: South America in the 1970s and '80s; middle eastern countries during the 1990s; and noticeably towards Islamic countries which might support global terrorism during the 2000s. All too often, aid serves to prolong bad government.[2]

2. To what extent **does aid actually reach developing countries**, much less the people? All too often, mismanagement, corruption and capital flight have wasted aid funds and even served to prop up dictatorial regimes in developing countries.[3]

3. To what extent **does aid actually increase the well-being and living standards of people in developing countries?** Critics quickly point out that there is little correlation between the amount of aid and growth or development. On the other hand, there are a number of economic arguments for aid. In helping developing countries to close the gap between saving and investment, developing countries can be aided by creating better fundamental conditions for growth and better living standards. This in turn can potentially lead to **mutually** beneficial trade and commerce arising between developed and developing countries. Additionally, one can argue that increased aid might help diminish conflicts and wars which have plagued the developing world for decades.

4. Has there been 'too much' aid? Donors increasingly suffer from aid '**saturation**' or '**fatigue**'. Aid sceptics argue that many countries simply do not have the capacity to absorb aid efficiently and that lack of infrastructure will cause aid monies to quickly run into diminishing returns. Adding to this, aid failures of the past might lead to a tightening of the purse strings as donor countries see squandered resources, corruption and failure to develop in many deserving areas which are competing for aid.[4]

5. How does aid affect countries ability to 'fend for themselves' and develop internally? There is rather strong evidence that high aid levels in lowest income countries **worsens productivity** and international competitiveness.[5]

---

1 WDR 2001, Chapter 11, page 190.
2 *Assessing aid – What works, what doesn't and why*, World Bank study 1998, page 49
3 An absolutely blistering account of aid mismanagement is to be found in a rather shocking book from the late 1980s; *Lords of poverty* by Graham Hancock. It is now freely available online and I actually do recommend reading a few chapters. Recommended…but severely depressing.

4 See for example HDR 2005, page 96 and
5 See for example *Aid, Dutch disease, and manufacturing*, Journal of Development Economics, 2009

Aid and development – a very tricky evaluation

Aid is of course a wide term. It covers a span between soft loans for long term investment in infrastructure to short term emergency relief for famines and natural disasters. Any discussion revolving around '*pro et con*' must of course take this into consideration.

Arguments for aid

There are a number of arguments in favour of aid which can be considered generally valid:

- **Aid and government spending:** Developing countries derive great benefits by being able to increase education, health and infrastructure and aid can help. As pointed out in Section 5.3 (*Institutional and political factors*) there is frequently a gap between the need to provide public and merit goods and the tax bases required to fulfil such needs. Aid can help fill this gap – in fact, aid represented 10% or more of all GDP in 21 African countries in 1997, and cutting this aid would have had most damaging effects on tax receipts and government spending.[6]

- **Aid and technology/capital:** Foreign currency can help developing countries to import the capital and technology necessary for economic growth. This is often coupled with direct technical and training assistance from the donor.

- **Aid and aggregate demand:** Aid can help create demand pull conditions and stimulate aggregate demand and investment. Aid to some of the poorest countries in the world can be up to one third of GDP and up to six times greater than government expenditure.[7]

Criticism of aid

"All too often, aid is driven more by politics than by need, undermining its effectiveness." Ban Ki-Moon, UN General Secretary

A number of criticisms have been put forward from both developed and less developed countries. Aid has a large number of vociferous critics,[8] and as this debate has been on-going since the Marshall Plan in the late 1940s, the list below is in no way complete.

- **Poor economic efficiency:** The most vocal critics claim that aid in effect is largely inefficient and therefore wasteful. Far too much aid does not have development per se in mind, but is the result of geo-political considerations aimed at supporting friendly countries and allies. Many observers have also been highly critical of the amounts of money squandered on large 'status' projects such as airports and government buildings, and also of the simple fact that too much money is lost to capital flight, corruption and bureaucracy. For example, in 1993, the World Bank estimated that official development aid **giving** amounted to $US56.7 billion, while ODA **received** was $US45.1 billion – the rest was apparently sucked up in high administrative and overhead costs of disbursing aid monies.[9] Another example is Zambia. The World Bank in 2001 estimated that if Zambia had used its aid efficiently between 1961 and 1994, the country would have a per capita GDP of $US20,000 rather than $US600.[10] Another issue is **tied aid** – criticised in Chapter 91 – which is vociferously (= loudly) condemned in the Human Development Report 2005 for its wastefulness. In fact, the UK banned all tied aid in 2002, Ireland and Norway have followed suit, and some 12 additional OECD countries have promised to follow.[11]

- **Zero – or even negative! – correlation between aid and growth in the poorest countries:** Some of the harshest criticisms – and hence also some of the more debated – levelled at aid look at the dismal performance of the poorest nations' growth rates in relation to aid received. In a rather blistering critique of aid's inability to foster economic growth, development economist

---

8   One of the staunchest opponents of aid, (Lord) Peter Bauer, argued that developing countries were pretty much invented by foreign aid. A most memorable (?) quote of his was that aid is "an excellent method for transferring money from poor people in rich countries to rich people in poor countries." Quoted from the Economist, *A voice for the poor*, May 2nd 2002

9   HDR 1993, pages 172 – 174 and 203; WDR 1993, pages 238 – 239

10 WDR 2001, page 192

11 See OECD online under *Untying aid*. See also See *Where does the money go?*, Brookings Global Economy and Development, Working Paper 21, June 2008, William Easterly and Tobias Pfutze, page 17

---

6   *Taxation and tax reforms in developing countries: Illustrations from sub-Saharan Africa*; Development Studies and Human Rights, C. Michelsen Institute 2003, page ix.

7   World Development Indicators 2005, table 6.10 and World Databank

DEVELOPMENT

William Easterly found that there was strong evidence of negative correlation between aid and growth.[12]
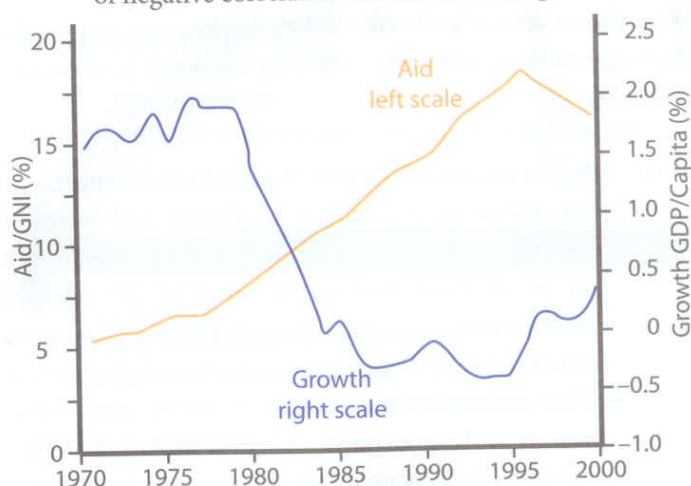
Figure 92.1 Growth and aid in Africa 1970 – 2000 (Source: http://news.bbc.co.uk/2/hi/science/nature/4209956.stm and Easterly, page 46)
As African aid rose, growth slowed. World Development Indicators Online

- **Aid diminishes local investment:** Aid can actually worsen the domestic production base in an economy by replacing domestic saving and investment (e.g. a form of 'crowding out'). Later research points out that aid has not closed the financing gap between saving and investment which was the claim earlier.[13] Additionally, aid in the form of money can increase imports and worsen the current account balance, while direct food aid has in several cases destroyed domestic market fundamentals by lowering the price of domestic crops and thus removing the incentives function of supply.

- **Control of monies:** The everlasting question is "Who decides where the money goes"? Donor countries, as stated above, frequently have strong political and economic motives in recipient countries. (For example, USD3 billion of US bilateral aid in 2012 went to Israel and 75% of all bilateral aid received by India came from the UK.[14] Get yourself a good history book and figure

out why.) In addition to this bias, donor countries will want some form of control over where the money goes. On the other hand, recipient countries might rightfully claim that they best know how to allocate funds. As an extension of this, there are numerous examples of how poor coordination between development agencies has resulted in mixed goals between donors and recipients, leading to inefficient use of funds and poor resource allocation.[15]

- **Aid doesn't reach the neediest:** Often it is countries which have relatively good institutions and infrastructure which receive the majority of aid, and most aid flows to a relatively limited area. It turns out that only a fraction of these flows reaches the poorest segments of the population. The UNDP estimates that official aid misses around 20% of the poorest people and that even non-government organisations specifically targeting the poorest communities might miss 5 – 10%.[16]

- **Aid helps to support evil regimes:** Another argument frequently put forward is that aid going to undemocratic and corrupt regimes aids bad governments in continuing to exercise power. William Easterly estimates that about a third of aid goes to countries classified as 'unfree' and that this proportion has been relatively stable over 30 years.[17]

- **What does it cost to get USD1 of aid to a country?!** High overhead costs – such as administration and salaries for those disbursing the funds – can make up a surprisingly significant proportion of total aid costs. For example, as a percentage of the total official development financing, overhead costs are estimated at 1% for Italy, 5% for the UK and 19% for the Netherlands. If we look instead at multilateral aid, the UNDP has overhead costs – hold on now! – of 129% of development financing. Removing other fixed costs, this means that the net contribution of every single international employee in the UNDP is the same as their salaries; one US dollar of aid costs USD2 in terms of salaries.[18]

---

12 William Easterly, *The White Man's Burden*, pages 42 – 55. Again I warn of drawing erroneous conclusions as to causality! So does Easterly, e.g. that the increase in aid is most likely due to the fall in growth rates. Yet he does allow for the distinct possibility that aid has not done much in the way of *stopping a decline* in per capita growth rates over the past 30 years in Africa.

13 WDR 2001, page 192

14 See US Congressional Research Service, docket 7-5700, March 12 2012 and *We don't want your aid*, The Telegraph, 4th Feb 2012

15 See for example WDR 2001, page 192

16 HDR 1993, page 96

17 See *Where does the money go?*, Brookings Global Economy and Development, Working Paper 21, June 2008, William Easterly and Tobias Pfutze, page 14

18 Ibid, page 19

Debt relief

Finally, there is the issue of **debt relief** as a form of aid. The debt relief granted to a number of heavily indebted poor countries (HIPCs) is an initiative aimed at alleviating poverty and the burden of debt in some of the world's poorest countries. Between 1989 and 1997, debt relief for 41 HIPCs amounted to $US33 billion.[19] An initial criticism of this procedure is that it gives no control whatsoever to countries writing off debt as to how funds which do not go towards debt servicing will be used. A recent criticism is also that debt write-offs during the 1990s have in fact had little or no effect on growth or development in the poorest countries.[20] In 2011 there are 40 HIPCs qualifying for a new round of debt relief and while there is a good deal of support globally for the effort, a number of commentators have expressed worries that writing off debt might in fact lead to wrongful behaviour in the future. Debt relief might create a situation known as 'moral hazard', which means that if the government in a developing country sees that there is a possibility of debts being written off, there is an incentive for over-borrowing.

Note from a famous economist: Jeffrey Sachs

Delving into the issue of whether aid 'works' or not is a bit like fly-fishing; the more one learns the less one understands. During the collection of sources and outlining of the previous version of this book, I was increasingly confused by the wide range of views on development aid by some very notable economists. The most acrimonious (= hostile) debate ranged between professor Jeffery Sachs (author of *The End of Poverty*) and William Easterly (author of *The White Man's Burden – Why the West's efforts to aid the rest have done so much ill and so little good*). Read the titles – guess which side these two stand on!

Easterly basically claims that there is little indication that the majority of aid money over the past 50 years – some USD2.3 trillion – shows little in the way of development for the recipient countries. In some 380 well-annotated pages, Easterly builds an argument that the well-intentioned 'planners' of development aid have wasted huge amounts of resources with little to show for it. In summa: aid is ineffective.

Sachs, who ran the UN Millennium Project between 2002 and 2006, is a strong advocate of massive scaling-up of aid. He very clearly points out that the aid money from the world amounted to USD12 per person per year to Africa – and that net aid to Africa from the largest economy in the world, USA, amounts to, wait for it, 6 cents per African. In other words, there is little to show for aid since there is little to show in aid.

Rather vaingloriously (= boastfully, or self-aggrandizing) I wrote to professor Sachs and asked him about the "Does aid work?" debate. He very kindly answered. Here is his response:

> As I am sure that you would gather, I find the debate a bit wearisome, because there is no such thing as 'aid' in general, but only aid for specific purposes. There are bad ways to do things and good ways to do things, including the provision of development assistance. The real question is "Can aid be properly managed (with high likelihood) to provide useful and important results?" not "Has 'aid' worked or not worked?" The latter question is too broad and in some ways silly.

> I would like to remind you that after a slashing attack on aid for more than 300 pages, Easterly actually writes the following:

> "Put the focus back where it belongs: get the poorest people in the world such obvious goods as vaccines, the antibiotics, the food supplements, the improved seeds, the fertilizer, the roads, the boreholes, the water pipes, the textbooks, and the nurses. This is not making the poor dependent on handouts; it is giving the poorest people the health, the nutrition, education, and other inputs that raise the payoff to their own efforts to better their lives." (White Man's Burden, p. 368-9)

> That's approximately what I've been recommending for years. Where's the real debate?

Aid vs. trade

> "Aid without trade is a lullaby – a song you sing to children to get them to sleep." Uganda's president, Yoweri Museveni[21]

Many economists claim that *fair and free trade* would in fact lessen the need for aid, and that there is a greater return on addressing trade issues such as barriers to trade in developed countries. A counter-position is that many developing countries are in much too weak a position to be able to compete on completely free trade terms and that domestic industries in developing countries would suffer from highly efficient multinational companies' oligopolistic – and even monopolistic

---

19 WDR 2001, page 203
20 WDR 2001, page 203

21 Quoted from the Economist, *An addictive lullaby*, January 15th 2004

DEVELOPMENT

– positions of power. Somewhere in between these two positions one finds a degree of consensus in that one does not exclude the other; aid can be channelled towards institutions, infrastructure and knowledge/technology transfers which both complement and enhance trade.

- Many developing countries are caught in producing and exporting **primary goods** which have low price and income elasticities of demand. Thus their terms of trade have been falling as the relative prices of these commodity exports have fallen continuously for decades.

- Many **barriers to trade** still remain in developed countries for precisely the goods which less developed countries have a comparative advantage in, primarily agricultural goods and textiles.

- While **free and fair trade** is generally advocated as a method for developing countries to increase income, there are risks involved in removing trade barriers in developing countries. Among these are, notably, the harmful effects on domestic production and employment as a result of highly efficient foreign competitors; loss of tariff revenue which makes up a large portion of government tax receipts; and a developing country with strong outward-orientation could be highly vulnerable to international business cycles.

The trade advocates say…

Proponents of trade as a major development driver point out that highly successful newly industrialised countries (NICs) have followed strongly export-oriented paths based on producing according to comparative advantages. The Asian Tiger economies, and more recently China and India are commonly referred to examples. A great many in-depth studies by the World Bank, OECD and the WTO show strong indicative evidence that successful integration into global trade flows is **strongly pro-growth and pro-developmental**. For example, the OECD estimates that a 75% cut in tariffs and subsidies in rich countries would raise LDC incomes by approximately USD23 billion – a GDP increase of 0.3% in Sub-Saharan Africa South Asia and Latin America. Furthermore, of the estimated USD97 billion gain in incomes due to lower tariffs for non-agricultural

goods, USD68 would go to developing countries.[22] Figure 92.2 illustrates that increasingly outward-orientated countries have not only outpaced other countries but are also pro-poor, in that the poorest 20% of the population enjoys the strongest relative income growth.
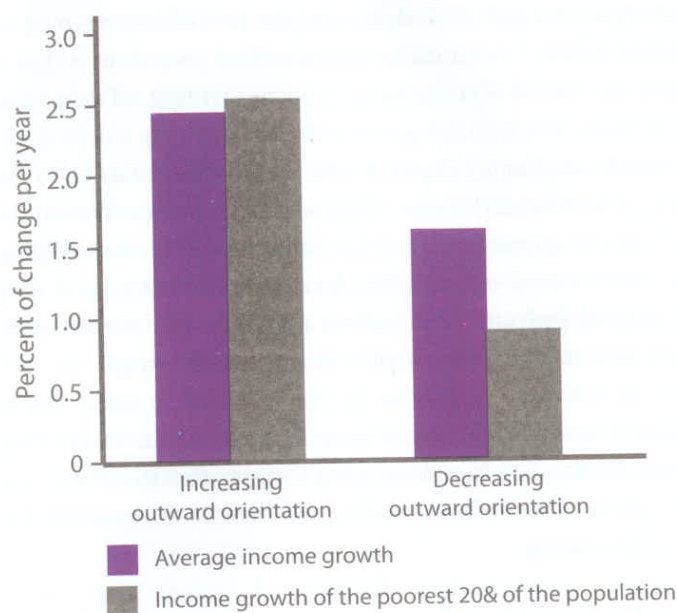
Figure 92.2 Export-orientation and pro-poor growth.
Source: Global economic prospects and the developing countries, World Bank 2001, page 44.

However …

As more *trade-cautious* commentators point out, income growth in fast growth countries is not evenly spread and has led to increasing **income inequality** – primarily between urban and rural citizens. There are also negative externalities such as very high **pollution** levels in some cases, China being a case in point. In addition, there are no guarantees that the success formulas of the NICs can be transferred to any and all other developing countries. Finally, there is very weak evidence that trade liberalisation has had any effect on **poverty reduction** in the **least** developed countries – in fact, there is more evidence that rapid and thorough trade liberalisation led to *increased* poverty in the short run.[23]

22 See for example http://www.oecd.org/document/33/0,3746, en_2649_37431_43440634_1_1_1_37431,00.html and Global economic prospects and the developing countries, World Bank 2001, chapter 2

23 *The least developed countries report 2002;* UNCTAD, pages 117 – 119

DEVELOPMENT

Good policies + good institutions + aid = growth (?)

Summary & revision

While there is no conclusive outcome in the aid debate, a number of indications point to the fact that aid indeed has had some notable successes in countries which have strong basic institutions and sound economic policies. This is not surprising and really only strengthens was has been previously said; development is a process involving multiple and simultaneous variables. Aid monies will only be effectively used when there is/are: a functioning financial systems, e.g. **banks**; sound **macro policies** resulting in price stability and balanced budgets; **aid agendas** based on needs rather than grandiose development schemes; **good governance** based on openness and accountability; efficient administrative institutions under the **rule of law** which minimise corruption and theft.[24]

1. The **aid debate** centres around questions such as the extent to which aid to LDCs is politically motivated; actually reaches those in need; and actually works.

2. **Arguments for aid** include:

   a. Aid can *boost government spending* on much-needed infrastructure and public services

   b. *Technology, knowledge* and *capital* can be transferred

   c. Aid can *stimulate aggregate demand*

3. **Criticism of aid** usually focuses on the questionable efficiency of aid:

   a. High *overhead costs* of development agencies mean that a good deal of money goes toward salaries and administration rather than the people in the recipient countries

   b. There is evidence that significant amounts of development funding to some countries over decades have in fact *not fostered growth*

   c. Aid can **crowd out** local investment

   d. Aid often does not *'filter down'* to the neediest people

   e. Aid has often helped *support evil regimes*

4. Many economists have pointed out that **free and fair trade** – e.g. where LDCs can compete based on their comparative advantage without competition-skewing trade barriers – is far more efficient than aid in the long run.

5. There is evidence that **aid is most effective** in countries which have *sound institutions* and political stability.

---

24 *Assessing aid – What works, what doesn't and why*, World Bank study 1998, pages 2 - 6; and the Economist, *An addictive lullaby*, Jan 15th 2004. (It should be noted that a follow up report using later years for data failed to show convincing results that aid indeed promoted growth and development in 'good policy environments'.)

93: Multilateral Development Assistance – IMF and the World Bank

Key concepts:

- IMF
- World Bank

> The main IMF prescription has been budgetary belt tightening for patients much too poor to own belts.
>
> Jeffrey Sachs (*The End of Poverty*, page 74)

The 45 nation Bretton Woods conference in July, 1944, created four institutions which were intended to increase stability in the post-war period:

- the fixed exchange rate system which became known as the Bretton Woods system (1947 – 1971);

- the General Agreement on Tariffs and Trade (GATT) which became the World Trade Organisation (WTO) in 1995;

- the International Monetary Fund (IMF) and

- the World Bank.

IMF

The **International Monetary Fund**, with headquarters in Washington DC, came into effect with 39 members in 1945 and is based on membership. Member countries pay what in effect is an annual membership due to the IMF. This is based on the size of each country's economy. The size of the membership dues control members' voting rights – unlike, for example, the WTO, where each country has equal voting rights. The US has by far the largest percentage of votes, 16.75%, followed by Japan (6.23%) and Germany (5.81%), and France (4.29%).[1]

The original purpose of the IMF

The membership dues created a joint-owned fund which could be used to stabilise currencies which had difficulties in keeping the fixed exchange rate created at Bretton Woods. In acting as an overseer of the fixed exchange system, the IMF had two primary functions, firstly to **monitor the balance of payments** in member countries to detect signs of fundamental disequilibrium and, secondly, to **intervene on behalf of countries** which showed serious balance of payments problems and downward pressure on their currencies.

Recall from Chapters 71 and 72 that a current account deficit puts downward pressure on the exchange rate so, in a *fixed* exchange rate system, the currency would in effect become overvalued. In order to restore credibility to the currency, a country would need to borrow foreign currency in order to bolster the value of its own currency. By allowing countries to borrow from the fund (*inflow* on financial account in the balance of payments) the IMF would enable countries to get through periods of current account deficits (*outflow* in current account) without having to devalue their currency. The IMF was basically to act as an **international lender of last resort** to the countries which were part of the Bretton Woods fixed exchange rate system.

Thus the IMF was to act as a guarantor of international exchange rate stability by enabling countries to borrow from a pool of funds collectively owned by members – a bit like an insurance company which is owned by a cooperative of insurance buyers. During the early 1970s, the Bretton Woods exchange rate system broke down and the following decade was marked by international instability and oil crises, stagflation, increasing current account deficits in developing countries and the debt crisis. All of this contributed to a considerable shift in the role of the IMF.

---

1 http://www.imf.org/external/np/sec/memdir/members.htm

The role of the IMF from the 1980s forward

Developing nations which had fallen into serious debt problems by the end of the 1980s also had severe government and current account deficits, together with high inflation and low growth. The IMF played a central role during the **debt crisis of the 1980s** (see Chapter 94) by acting as go-between for deeply indebted countries and creditors in developed countries. The IMF enabled indebted countries which were unable to service loans to *reschedule* their debts, i.e. take on further loans under longer time periods to meet current loan payments. Commonly, the creditor nation or commercial bank would work out an agreement with the debtor nation and the IMF. No country was willing to default on IMF loans as this would pretty much lock the country out of future loans and IMF assistance, so the rescheduling was made possible by the IMF in its capacity as guarantor against default (= non-payment of loans). In essence, having the IMF involved in the process lent creditor countries credibility, enabling future loans from banks and international institutions. The IMF did not indulge in 'free lending' in any way – on the contrary, countries which needed to borrow from the fund were subject to a number of very strictly imposed conditions by the IMF. This is the (in-) famous **IMF conditionality**, where a debtor would have to implement a 'package' of macroeconomic policies in order to meet IMF approval and the cooperation of the creditor bank.

The objective of the IMF is still focused on upholding currency stability/convertibility, and the supervision of policies to influence balance of payments problems. And while the IMF is adamant that it is **not an aid agency or development bank**, during the 1980s and '90s the it continued to act as a major creditor and greatly extended the lending operation to developing countries. The IMF now consists of 188 members and, as one the world's largest creditors to developing nations, has increasingly adopted a role in development in addition to the role of international monetary/financial stability and cooperation. It is in this capacity that such harsh criticism has been directed at the IMF in later years.

IMF stabilistion policies (structural adjusment policies)

A country which needs to reschedule debt and/or take on new loans from the IMF will meet a number of strict conditions imposed on the debtor nation. IMF loan conditionality comes in the form of so-called **stabilisation policies** (or **structural adjustment policies, SAPs**[2]), aimed at correcting fundamental macroeconomic problems. Generally, a stabilisation package imposed by the IMF aims to resolve balance of payments difficulties, contribute to sustainable growth and thus allow the debtor nation to repay the loan. These policies are strongly *market-orientated* and typically contain most of the following:

- Forms of currency market intervention, such as foreign exchange controls and intentionally overvalued exchange rates must be eradicated, and a freely convertible currency must be established. This facilitates increased exports and decreased imports, which in turn improves the **balance of payments**.

- **Inflation** must be combated, which often involves raising interest rates and decreasing government spending.

- The decrease in government spending is often complemented by an increase in taxation in order to balance the **government budget**.

- Limits on foreign direct investment must be curtailed to **increase FDI** inflows.

- A number of **supply-side policies** are often demanded by the IMF, such as the removal of minimum wages, doing away with minimum price schemes and other market distortions, cutting of subsidies on many goods, and privatising state owned enterprises.

- While it is not put in the form of a demand, the IMF encourages the removal of **trade barriers**.

The harshness of these policies has been defended by stating that sooner or later they will have to be done, and the later such policies are implemented, the more severe they will be for debtors. Countries which come to the IMF for assistance are in deep financial trouble already and will face fiscal hardship any which way one looks at it. The IMF provides loans which enable the country to adjust the economy quicker and with less pain than otherwise would be the case.

---

2 Now re-branded *Poverty Reduction Strategy Papers*, PRSPs, that LDCs must submit in order to qualify for IMF loans.

DEVELOPMENT

Criticism of the IMF

The stabilisation policies above serve as *indicators* of economic improvements in an economy for other potential creditors, e.g. commercial banks, enabling a country to get further loans outside the IMF. In other words, financial institutions see IMF involvement as a 'seal of approval' for loans, giving the IMF in fact a far more powerful role than solely a lender. As one consequence of this power, few organisations of latter days have come under such harsh scrutiny and disapproval as the IMF. The criticism has come from within and outside the IMF; from left and right; and from developing and less developed countries. While it is clearly impossible to include an in-depth study of all of these points here, the following four captions cover the main points of discontent. The first two represent a 'grassroots/anti-globalisation/anti-capitalism' view and the last two a 'neo-liberal/ capitalist/conservative' point of view.

- **Stabilisation policies are anti-developmental:** The harshest criticism has been whether the 'medicine' prescribed by the IMF does not, in fact, make the patient worse off. The harsh demands of stabilisation policies will have serious effects on development: Trade liberalisation destroys domestic markets and leads to unemployment; high interest rates stifle investment; cuts in government spending have devastating effects on much-needed education, infrastructure and welfare programs; reduction of food subsidies will affect the poorest groups the most; and increased inward FDI and market deregulation often creates monopolies due to poor competition rules and oversight.

- **The IMF is non-democratic and run according to rich country interests:** The Managing Director of the IMF is always from Europe and the richest 7 countries (the G7) can, in effect, block all proposals which require a qualified majority vote of 75%. The US alone has more votes than all of Africa together.[3] The harshness of stabilisation policies is always inflicted on poor countries and never on rich countries, and policies which require developing countries to lower trade barriers will serve the interests of developed countries rather than developing countries.

- **IMF lending leads to moral hazard:** The IMF crisis loans are simply re-worded 'bailouts' which serve to perpetuate bad loans and bad debtor countries. When countries realise that poor debt service can lead to further loans to reschedule debt and/or possible write-

off of debt, there is less incentive to behave in a morally responsible manner. Thus there is moral hazard.

- **IMF lending works poorly – if at all:** In far too many cases, IMF loans show poor results. Debtors have often failed to implement necessary reforms so the taxpayers of large IMF countries foot the bill. Brazil has received over $US53 billion in IMF loans since 1958, which has done little to open up the economy and implement market reforms. Argentina, the third largest IMF debtor, has implemented limited structural reform, and has had little increase in real per capita GDP in the last 30 years.[4]

The World Bank

The World Bank is often called a 'sister organisation' to the IMF. It was created at the same time and is housed in Washington. It too has 184 member countries whose voting rights are contingent upon economic size. Contrary to the IMF, the World Bank is indeed an outright development organisation created initially to deal with the reconstruction of war-torn Europe. This is indicated by the Bank's original name; the **International Bank for Reconstruction and Development, IBRD**. Together with the **International Development Association, IDA**, which was created in 1960, we have the grouping commonly referred to as the **World Bank**.

The overriding purpose of the World Bank is to create the groundwork for social and economic development by providing funds in the form of loans. It is, in reality, an investment bank, acting as go-between for investors and loan recipients, and is thus one of the world's largest borrowers on capital markets. The World Bank is owned by the 187 member nations and has offices in over 100 countries. The two institutions of the World Bank have the same general goals but have slightly different responsibilities.

IBRD: During the post-WWII period, the IBRD was focused on reconstructing the ruined economies by financing infrastructure. During the 1950s the IBRD became increasingly involved in helping developing countries with finance by diverting deposits from developed countries to developing countries via loans. The IBRD sells bonds on the world market and then re-lends the money to developing countries. The loans from the IBRD are on commercial terms or near-commercial

3 http://www.imf.org/external/np/sec/memdir/members.htm

4 Report from the Heritage Foundation, a US liberal think-tank, at http://www.heritage.org/Research/InternationalOrganizations/bg1689.cfm

DEVELOPMENT

terms, and there is strong emphasis on the debtors ability to service loans. In fact, the IBRD has run at a profit since 1948. Thus, in a similar fashion to the IMF, the IBRD has, since the early 1980s, demanded that debtor countries implement **structural adjustment programmes**. These are primarily open-market reforms and the removal of trade barriers as itemised under IMF stabilisation policies above.

IDA: The IDA is the **concessional** lending arm of the World Bank, and lends only to the poorest countries (with a GNP per capita mostly below $US865) which would not be able to borrow at IBRD rates or meet the demands of adjustment programmes. The IDA is basically a provider of *soft loans*, i.e. loans at less than market rates and with very favourable terms of debt servicing. Commonly, IDA loans are interest free, and have longer payback periods – often 35 to 40 years.[5] In contrast to the IBRD, the IDA receives most of its funding via donor contributions from member countries.

In addition to financing loans, the World Bank has increasingly expanded operations to a wider field of development than pure economic development. It is one of the major sources of development statistics and research, and the yearly **World Development Report** is highly regarded as one of the most important compilations of development statistics. The Bank also provides development assistance, technical support and education in over 100 countries via local branch offices.

...and the critics say...

The criticism of the World Bank follows the same lines as for the IMF; the consequences of **conditionality** in borrowing from the Bank are very negative for the poorest parts of developing countries. Fiscal austerity (severity) imposed by World Bank officials has been shown to have clear anti-developmental effects on the poor. In fact, the UNDP's Human Development Report has shown clearly that the Human Development Index (HDI) has fallen in a number of countries which have implemented structural adjustment programmes in order to qualify for World Bank loans.[6]

Harsher – and perhaps more ideologically tainted – criticism comes from many NGOs operating in development. It is pointed out that the World Bank is clearly run by the richest countries, with a strong American presence. For example, every president of the World Bank Group[7] has been an American,

and the US has almost 16% of the voting rights while China and India – accounting for one third of the world's population –together have 7.33%.[8] Critics on the neo-liberal, free market side claim that a good third and possibly half of all World Bank projects are failures.

Perhaps the heaviest criticism comes from **Joseph Stiglitz**, ex-chief economist at the World Bank and Nobel Laureate, who has harshly accused the World Bank for wrongful policies and basic incompetence over many years. His main point is that the structural adjustment policies have been so obviously damaging for many years, but have not led to fundamental shifts in World Bank methodology and practice[9]. Stiglitz has pointed out that countries which opted out of World Bank loans due to the harshness of conditionality, e.g. Malaysia in the late 1990s, have frequently fared far better than countries which took the loans and subsequently applied the market liberalisation policies prescribed by World Bank structural adjustment demands.

---

5   See an excellent overview of both the World Bank and the IMF at http://www.imf.org/external/pubs/ft/exrp/differ/differ.htm

6   HDR 1990 in *Todaro*, page 631

7   There are **five** institutions in what is called the **World Bank Group:** The IBRD, IDA, the International Finance Corporation, the Multilateral Investment Guarantee Agency, and the International Centre for Settlement of Investment Disputes.

8   See World Bank at http://web.worldbank.org/WBSITE/ EXTERNAL/EXTABOUTUS

9   *A plague of finance*, the Economist September 27th 2001

Summary & revision

1. The International Monetary Fund, **IMF**, is a multilateral organisation which aims to increase international currency stability and convertibility, aid countries with balance of payments problems and help nations with credit problems restructure debts.

2. **Structural adjustment policies** (SAPs) by the IMF and World Bank are market-orientated policies put forward as conditions for receiving loans. Common elements of SAPs are:

    a. Freely floating exchange rates

    b. Government focus on curbing inflation

    c. Balanced budget

    d. Openness to FDI and trade

    e. Supply-side measures such as privatisation and abandonment of market distorting subsidies for foodstuffs

3. The IMF has been **severely criticised**:

    a. SAPs can be highly anti-developmental when government spending on education and health care is cut to balance the budget and market deregulation can cause large scale unemployment

    b. The IMF is run primarily by the richest countries

    c. IMF lending can lead to moral hazard

    d. The IMF has a poor track record in terms of helping indebted countries deal with debt and attain solid growth over time

4. The **World Bank Group** is the world's largest development organisation. It offers loans/financing and offers policy and research advice to developing nations.

5. The World Bank has also been *criticised for the harsh conditionality* imposed on debtor countries.

94. The Role and Consequences of Foreign Debt

4.7

Key concepts:

- Reasons for indebtedness
- A short history of third world debt
- The long run cost of debt
- Debt relief – heavily indebted poor countries (HIPC)

"There has never been any successful austerity program in any large country," Nobel laureate Joseph Stiglitz in 2012 on the subject of Europe which is "headed to a suicide" as euro-area leaders focus on austerity.[1]

Development economists stress the importance of developing countries having access to financial markets in order to increase investment. In particular, to have access to *foreign* capital, since domestic savings and institutions are often unable to provide adequate funds. Unfortunately, many of the poorest countries are still suffering from what became known as the **debt crisis** of the 1980s, and servicing *foreign debt* will have to be done using hard currencies such as the dollar or yen, which mostly must be obtained via exports. Adding to the problems of finding investment sources is the fact that large amounts of money have been placed in foreign banks rather than funding domestic investment and consumption, so-called capital flight.

Reasons for indebtedness

There is no major difference between property owners borrowing money and countries borrowing it – capital is needed to improve output and the productive value of the land. When developing nations borrow on foreign capital markets it is often due to a very low level of domestic savings as mentioned previously. Total debt in developing countries went from $US68 billion in 1970 to approximately $US2.3 trillion in 2002 – an increase of over 3,000%.[2] Within this there is a group of most **heavily indebted poor countries** (the 39 **HIPC** countries) which accounted for close to $US190 billion of debt in 1998/99 – which sounds proportionately little until one realises that this represented over 400% of these countries' total exports in 1999.[3] How did so many developing countries end up with such amounts of clearly unsustainable debt?

A short history of third world debt

The debt crisis in developing countries became all too clear in the 1980s and had its origin in the quadrupling of the price of oil in 1973/74 and resulting supply shocks around the world. Oil is paid for in US dollars and the massive increase in revenues for OPEC and other oil producing nations resulting from higher oil prices flowed into commercial banks in developed countries. The deposits of petrodollars in commercial banks – primarily

---

1   www.moneynews.com 27th April 2012

2   IMF at http://www.imf.org/external/np/exr/ib/2000/092300. htm; Intergovernmental Group of 24 at http://www.g24.org/ losertab.pdf; and Todaro, page 608

3   *World Economic Outlook*, IMF 2000, chapter 4 page 141

in the US and Europe – increased almost tenfold between 1973 and 1974. This increase in loanable funds must be seen in the light of two additional factors:

1. The **Bretton Woods system** of fixed exchange rates had broken down. Countries were thus free to set interest rates to pursue domestic macro policies in the process.

2. The oil crisis resulted in a world-wide **recession in 1974/75**. The recession was unusual as it was *inflationary* due to the supply-side shock of increased oil prices. This unprecedented *stagflation* caused world wide declines in output and thus falling demand for commodities.

There was a large increase in loanable funds and low demand for loans in developed countries. Commercial banks looked for alternative borrowers and the scene was set for a surge in lending to developing countries – countries are low risk as they do not go bankrupt like firms. In addition, *real* interest rates were very low due to high inflation and also due to the fact that developing countries relaxed credit in order to stimulate demand. Accordingly, debt in developing countries increased by over 22% annually between 1975 and 1980, going from a total of $US180 billion to $US500 billion.[4]

When oil prices once again rose drastically in 1979/80, there was a shift in monetary and fiscal policies in the developed world, primarily in the US. Newly-elected American president Ronald Reagan lowered taxes and ran up large budget deficits which were covered by loans – which of course drove up interest rates and the exchange rate for the US dollar. Interest rates skyrocketed both in the US – doubling in real terms between 1979 and 1982 – and other developed nations also due to deflationary policies aimed at coming to grips with high inflation. The results were nothing short of disastrous for developing countries:

- As a large proportion of developing countries' debt was *denominated in US dollars*. Debt servicing (i.e. paying off the loan – amortising – and paying the interest) became ever more burdensome. More export revenue was needed to pay off debt. Figure 94.1 shows the appreciation of the US dollar between 1979 and 1985.
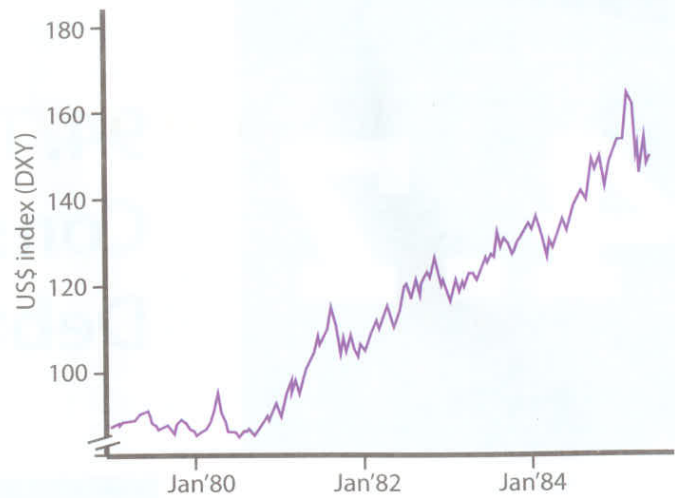
Figure 94.1 USD exchange rate (index), 1979 – 1985

- The loans in commercial banks were at market rates and often variable. The increase in interest rates thus increased the *cost of the loans*.

- A large portion of the borrowed funds going to developing countries bounced right back to bank accounts in developed countries. This is known as *capital flight*, and is looked at below.

- Continuously *falling commodity prices* hurt exports and many developing countries went into recession, which further decreased their ability to service loans. When Mexico defaulted (= declared itself unable to amortise and pay interest) in 1982, the debt crisis was established as fact.

The long run cost of debt

"…foreign aid is necessary to enable underdeveloped countries to service the subsidized loans…[taken] under earlier foreign aid agreements." Lord Peter Bauer

Again, it must be pointed out that taking on debt is not necessarily a bad thing for developing countries. Borrowed funds are a way of filling the finance gap, i.e. the gap between domestic savings and investment needs. Foreign loans are one method of providing this finance, but when the debt is beyond the country's capacity to service the debt in the long run, the results can be devastating. The 'double-whammy' outlined previously, where rising interest rates and an appreciating US dollar hit debtor LDCs, resulted in impossible debt burdens for some 50 developing nations in the mid-1980s.

4   *World Economic Outlook*, IMF 2000, chapter 4 page 139 and http://www.imf.org/external/np/exr/ib/2000/092300.htm
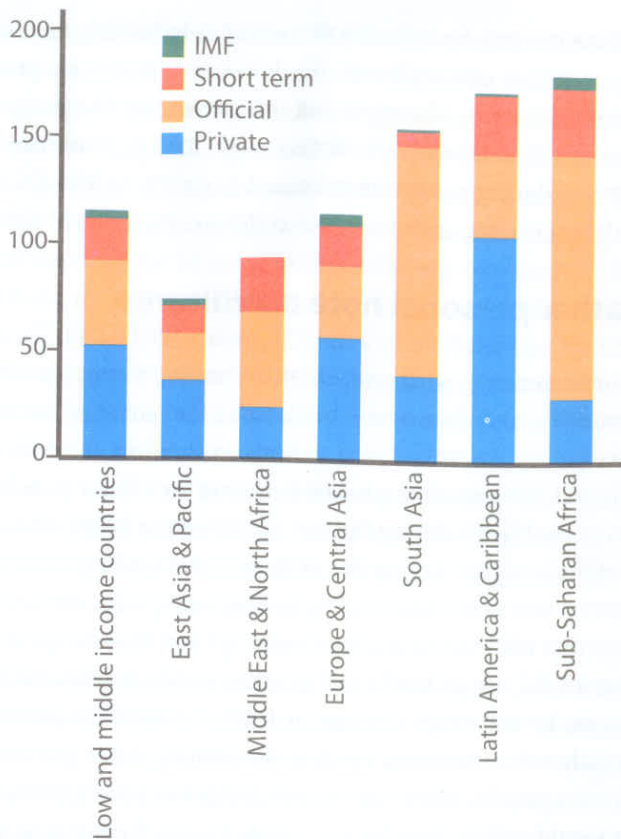
Figure 94.2 Total foreign debt in developing countries in 2000 as a percentage of exports

Note: LMICs = Low and middle income countries

(Source: *Global Economic Prospects and the developing countries*, World Bank 2002, page 19)

Debt has shown to have some major anti-developmental effects:

- **Diversion of funds:** Since the largest proportion of loans in developing countries have been taken by governments, *official debt* in Figure 94.2,5 large **debts divert much-needed funds for roads, health care and education** to debt servicing (e.g. payments on the loan itself plus interest). When debt servicing amounts to 20% of total government expenditure while 13% is spent on education and 6% on health, the opportunity costs of large debt become obvious. This is in fact the case in some 30 least developed countries.[6] In addition, aid monies from abroad supposedly going towards development – and often immediate relief – are all too frequently used to pay off debt. Africa is estimated to have spent four times more on debt servicing than on health care during the 1990s.

---

5 While debt is often put in terms of the percentage of total debt to GDP, it is more meaningful to put debt ratios in terms of debt as a percentage of debt to exports. This is simply due to the fact that many developing countries will have to earn foreign currencies to pay off external debt.

6 See for example *IMF World Economic Outlook* 2000, page 143

- **Further debt and current account deficits:** Many developing countries found themselves locked into a 'perpetual debt mechanism' in the 1980s, when governments took on additional debt to service old debts. Many governments simply hoped that additional debt would buy some time for exports to increase and thereby be able to service additional loans. The loans which could have been used to fill the investment gap were often squandered or simply became part of capital flight. The loan inflows in financial accounts of course meant large current account deficits.

- **Harsh domestic macro policies:** When countries get a reputation as '**bad borrowers**' they will find it increasingly difficult to get loans from private sources – and when they are able to procure loans they will pay a higher interest rate as a risk premium. The other alternative has often been to **refinance loans** (renegotiate loans at lower interest and/or longer amortisation periods) with the IMF as a guarantee of 'good fiscal and monetary conduct' on the part of the developing country. 'Good conduct' often meant that the debtor LDC would have to submit to a number of strict policy guidelines of fiscal austerity (decrease government spending and/or increase taxes) and monetary tightening (increase interest rates). The **severe conditionality** imposed by such **IMF stabilisation policies** (SAPs) have been criticised for the anti-developmental effects of decreased government spending and the deflationary effects of higher interest rates.

POP QUIZ 94.1

1. What is 'debt servicing'?

2. Why is debt so frequently put into relative terms, e.g. as a ratio of debt to national income or debt to exports?

3. Why might a debt burden put as 'percentage of exports' show a better picture of a debtor's ability to service the debt? Burkina Faso as an example here would offer a very good explanation.

4. Why did LDCs see debt ratios rise so drastically during the 1980s?

5. Using a diagram, explain why IMF/World Bank stabilisation policies/SAPs might in fact worsen the situation (at least initially) for LDCs caught in high foreign debt.

6. A 'standard LDC' in the 1980s was one caught in high US dollar denominated debt receiving a large proportion of export revenue from exporting commodities. Explain how this situation, together with an appreciating US dollar and higher US interest rates affected the current account of said 'standard LDC'.

Debt relief – heavily indebted poor countries (HIPC)

As brought up in Chapter 92, the 1996 HIPC initiative by the IMF and World Bank for cancelling debts in the poorest nations was aimed at relieving the poorest countries of debilitating foreign debts. The HIPC was reviewed and updated in 2005 to match the Millennium Development Goals (MDGs, see Chapter 80) and by 2006 some USD19 billion in debt had been cancelled for 18 countries – effectively halving their debt. Of 40 countries eligible for the HIPC initiative, 32 will in all likelihood receive full debt cancellation from debtors.[7]

There are some basic demands on countries applying for HIPC relief:

1. Be eligible to borrow from the World Bank's IDA (see Chapter 93)

2. Have a level of debt considered unsustainable

3. Must have a good track record of reform and sound policies via IMF and World Bank programs

4. Have a Poverty Reduction Strategy Paper (PRSP – see Chapter 93)

In order to finally qualify for debt relief under the HIPC initiative, the country must maintain its status as 'good debt performer' and implement reforms outlined in the PSRP.

The aims of the HIPC initiative, as outlined in the MDGs, are to enable poor nations to increase spending on programs designed to reduce poverty. By reducing the overall debt burden, indebted poor countries can reallocate funds towards social programs within health care, education and infrastructure. The program has met with a measure of success – by 2006 expenditure on poverty-reducing programs increased from 7% to 9% of GDP and debt servicing costs were reduced from 4% to 2% of GDP.[8]

A rather personal note on vultures

It is quite amazing what one learns by looking things up. I was astounded – and dismayed – to discover that some of the very thinking that got the poorest nations in the debt mess in the 1970s and '80s was acting to work against the efforts to reduce debts in the HIPCs during the first decade of the 2000s. You see, one of the 'insights' during the 1970s was that lending money to countries was safer than lending to businesses – for the simple reason that while firms could go bankrupt and thus escape debt, countries did not go bankrupt! In other words, a debt could be taken on by a corrupt dictator in 1978 and used for personal gain rather than building roads and hospitals. After numerous civil wars, revolts, usurpations and changes in government, the debt would still be held by the country – e.g. by the sons and daughters of the 'original debt generation' now in government 30 years later.

Some 75% of the total debt in HIPC initiative countries is held by the IMF, World Bank, African Development Bank, Inter-American Development Bank and other multilateral institutions. These have been most forthcoming in cancelling debt. Then there are other financial institutions holding the remaining 25% of the debt. They have been…less forthcoming. At the heart of the matter is the fact that the HIPC initiative is *voluntary* – i.e. the creditors basically need to be persuaded to cancel debts and the IMF/World Bank have been quite successful in this. However, and this is a big 'however'; it turns out that billions of dollars worth of debt had been bought by 'vulture funds' – private financial institutions that would buy debt cheaply and then *sue debtor nations for full payment*.

'Is that debtor over there coughing blood?!"

7   IMF factsheet, *Debt relief under the HIPC initiative*, September 2011

8   IMF at http://www.imf.org/external/np/exr/facts/hipc.htm

Say that a private bank holds USD80 million worth of debt from a poor country. The bank sees the writing on the wall during the latter part of the 1990s as increasing pressure from NGOs and development agencies demand that unfair debts be cancelled. The vultures circle. An unscrupulous fund manager from Vulture Fund Inc. now offers to remove this embarrassing IOU from the bank for 10 cents to the dollar, e.g. buy the IOU from the bank for USD8 million. Rather than cancel debts, the bank can sell the USD80 million IOU to Vulture Fund Inc. for USD8 million. Better to receive 10% than nothing – and not face the shame of refusing to cancel debt. Now the IOU is in the hands of Vulture Fund Inc. and, having paid USD8 million to buy the IOU, they want their ~~pound of flesh~~ rate of return! Now Vulture Fund Inc. gets on the phone to the troop of litigation lawyers held on retainer for just such an occasion and sues for the full value of the debt plus interest.

Seldom does one see such bald-faced cynicism. Quite frankly, I didn't believe what I was reading until I saw IMF documents explaining the situation. By now you've read a few pages of my writing and probably suspect that I used my ample artistic license and simply made up the very nasty but apt term 'vulture fund'. No, I didn't.

The IMF did!

Summary & revision

1. The **background of the 1980s debt crisis** is to be found in the increased liquidity arising from the 1970s oil crises – large amount of 'petro dollars' available for loans mean many LDCs borrowed heavily.

2. Many LDCs saw debt spiral out of control during the 1980s as *interest rates soared* and the US dollar (in which a lot of foreign debt was denominated) appreciated.

3. High foreign debt has **severe consequences for LDCs:**

   a. *Funds are diverted* from much-needed social goods

   b. *Current account* deficits

   c. IMF and World Bank structural adjustment policies (*SAPs*) for indebted poor countries often had *debilitating effects on the domestic economy*

4. The Heavily Indebted Poor Country (HIPC) initiative by the IMF and World Bank in 1996 has managed to reduce debt burdens for some of the poorest countries in the world.

DEVELOPMENT

4.8

95. Market Orientated Policies

Key concepts:

- The policy shift of the 1980s
- Benefits of market-oriented policies
  - Trade
  - Privatisation
  - FDI
- Negative outcomes
  - Market failure
  - Dual economy
  - Income inequality

The period between the end of World War II and the beginning of the 1980s was marked by heavy **government intervention** in virtually all developing countries. During the course of the 1980s there was a general shift towards market strategies based on neo-classical thinking. Much of this gained momentum during the period of accelerating globalisation during the 1990s.

The policy shift of the 1980s

During the first 30 years after the Second World War, scepticism towards the market mechanism and the belief that government intervention could be used to correct market failures and provide growth led to extensive state planning and control of production assets. This was also combined in many cases with *import-substitution strategies* and *nationalised industries*. Other common elements of interventionism were *central pricing policies* on many goods; *limited financial freedom* in order to ensure that government sponsored industries received the brunt of financing; an *increasing government sector*; and *overvalued exchange rates* to enable cheaper imports of capital

goods. A good deal of the government intervention was rather heavy-handed, for example land was sometimes forcibly taken by government and then allocated according to a central plan; agricultural produce prices were often set centrally and farmers were forced to sell at lower-than-market rates; and rural economies were often distinctly disadvantaged in taxes and subsidies in comparison with urban areas.

It looked good in theory, but was generally a failure. By the end of the 1970s it was becoming increasingly obvious that countries which had implemented market-oriented strategies were growing at a far higher rate. Several key reasons for the limited success of interventionist methods prior to the 1980s stand out:[1]

- Perhaps the main reason is that government intervention requires a **framework of administration and control** that simply didn't exist in most of the developing countries at that particular stage. Inefficient bureaucracies, poorly educated staff, and very little in the way of monitoring progress or control do not make large scale and detailed development planning programs readily applicable.

- As governments increased in size and tax revenues did not increase proportionately, large and increasing **government deficits** became the norm. This helped fuel replacement funding of government spending via loans from abroad. (See *Debt crisis* in Chapter 94.)

---

1   For a rather scathing review of interventionist planning in development, see *The White Man's Burden*, by William Eastery, Chapter one.

- The increase in foreign loans together with imports financed by an over-valued currency caused significant **current account deficits** in most developing countries.

Benefits of market-oriented policies

The three points above should be most recognisable to you; this is a 'freeing the market' process which is very similar to supply-side policies. The beginning of the 1980s saw the emergence of neo-classical theory, primarily in Reagan's America and Thatcher's Britain. Growth had been shown to be far stronger in economies which had a higher degree of market forces which were not controlled or limited by government intervention. The seeming failure of interventionist policies in developing countries helped shift many governments towards **market liberalisation** policies and increased outward-orientation. Some or all of the following market-oriented reforms were implemented to some degree in developing countries during the 1980s and '90s:

Macro economic balance

Many of the policies implemented were part of a general macroeconomic **structural adjustment programme**. This was a process recommended by the World Bank and the International Monetary Fund aimed at restoring general macro equilibrium in the economy.[2] (And received a large amount of criticism during the 1980s and '90s.) The goal of structural adjustment was to restore *balance of payments* and government *budget balances*, reduce *inflation*, stabilise *exchange rates*, and generally create conditions for long term economic growth. This meant that the public sector in many developing countries was considerably reduced while inefficient state enterprises were privatised. This was often accompanied by a removal – or at least reduction – in central price-setting which had distorted market prices.

Privatisation

Privatization of land and capital was a key element in forming economies based on market structures. In China, for example, what was possibly the largest single push of privatisation put most of the land in private hands by the beginning of the 1980s. The results were remarkable; agricultural output increased by an average of close to 8% per year enabling a country which 20 years earlier saw millions of famine fatalities now had an

exportable excess of produce. The increase in income brought some 500 million people in China out of extreme poverty – which the World Bank refers to as the largest and fastest decrease in poverty in history.

FDI and trade

Encouragement of **foreign direct investment and trade** has been a marked element in the market process in many countries. For example, Vietnam's '*doi moi*' (renewal) policy starting in 1988 had a central (planned!) theme of attracting foreign companies. When President Clinton did a little '*doi moi*' of his own in 1994 by lifting the trade embargo the US had in place with Vietnam, there was a surge of American FDI from the US to Vietnam. The capital stock in Vietnam rose from 3.6% of GDP in 1990 to 55.6% by 1999.[3] Poverty reduction was such that while three quarters of the population lived on less than one US dollar a day[4] in 1988, the figure was 37% by 1998[5].

Did the market policies work in developing countries? Sorry, the answer is 'yes – and no'. Yes, market-led strategies have increased incomes and lowered poverty levels (generally) in some cases, most notably in the Tiger Economies, China, India, Malaysia, Mexico and Vietnam. In other cases, notably Argentina, Brazil, and many of the transition economies in former Eastern Europe, the results have been mixed – as have the degree of market orientation, according to some scholars – and have not met with the success seen primarily in Asia. A third group must pretty much be characterised as dismally failing, and the majority of these countries are to be found in Sub-Saharan Africa.

Negative outcomes

It bears repeating that a number of the 'non-interventionist' market-oriented economies mentioned above were in fact to a lesser or greater extent steered by *government micro management policies*. Entire industries in Korea, Singapore and Malaysia all had very powerful government institutional backing and support during the 1960s and '70s. It was made abundantly clear – via subsidies, tax breaks, protectionism

---

2 The strongly market-oriented view of development policy by the World Bank and IMF during the 1980's and '90s became known the '**Washington Consensus**' – as both of these institutions are based in Washington.

3 *Foreign direct investment in Vietnam: an overview*, Freeman, Nick J.; Visiting Professor, National Economics University Business School, Hanoi, September 2002, page 17

4 *Norborg*, page 41

5 *Legrain*, page 54

and what not – to certain key businesses that the focus was on cooperation between firms and a focus on exports rather that on domestic competition and import-substitution. There have also been strict limits to foreign direct investment inflows in some cases, notably Korea and Taiwan.

This illustrates how difficult it is to draw general conclusions as to the *general* superiority of market-led and outward-oriented strategies. Some economists view the East Asian 'miracle' as proof of the validity of the free-market argument, while others would view the success simply as good open market policies combined with skilful government intervention. Few commentators deny all of the negative outcomes listed below.

Sustainability issues

Providing growth to growing populations while maintaining future output potential and maintaining our environment and natural resources will be quite a balancing act. Sustainability will depend on the 'mix' of our use of resources and human/fixed capital, and on the speed with which we are able to increase our *efficiency* in use of resources. Two to three billion more people will join us in the next 50 years and output will quadruple to $140 trillion. The following questions are at the heart of the **sustainability dilemma**:

- Since the billion people in wealthy developed countries consume 70% of the world's resources and are a major contributor to environmental degradation, what will happen if developing countries follow the same development path as developed countries?

- The poorest billion people also have severe environmental impact as they often have no choice but to use limited resources in an unsustainable manner. The solution here is poverty alleviation – but the problem of sustainability in growth remains.

- Can efficiency in the use of renewable resources increase at a pace that will keep up with growth of incomes and population? What about non-renewable resources; can they be used at sustainable rates and/or can we find viable substitutes?

- Can economic growth and prosperity in both developing and developed regions be sustained given the considerable externalities visible in terms of urban overload and inner city congestion?

Enjoy the trees in the background…while you can. (Indonesian rainforests being razed to make room for palm oil plantations.)

No, I haven't provided any answers – and *nobody really can*, as yet. I direct this statement towards the numerous environmental organisations which have been predicting doom and gloom for some 40 odd years now. Hard-line environmental groups claim that we will have to fundamentally change our consumption and production habits; travel less, trade less, and simply adjust our mouths to a given food parcel. Optimists and technophiles set high hopes on the ingenuity of people to ultimately come up with solutions: the cry of the doomsayers in the 1960s and 1970s that we would be fighting for air, water, food and oil by the year 2000 has been proven false.[6]

The entire issue of sustainability is often put into a 'development' context, which in fact becomes something of a contradiction in terms. Externalities such as pollution, global warming and depleted fish stocks are common to *all* countries. In summing up the concerns of a sustainable development path in the future, The World Bank outlines several 'responsibilities' for:[7]

Developing countries: Strengthen **institutions** such as governance, rule of law and banks to enable a virtuous cycle of income → saving → investment. There must be a strong focus on **social investment** such as infrastructure and education, and it must be as broadly accessible to all as possible. All **institutional reform** must be transparent and those in charge must be openly accountable.

Developed countries: The key for developed nations is to enable poverty reduction by providing invaluable transfers of aid, knowledge, capital, and technology in order that developing countries may increase the speed of development. In lowering

---

6 Read anything by Rachel Carson, Lester Brown or Paul Ehrlich. Then refresh yourself with anything by Björn Lomborg, Ronald Bailey or Indur Goklany.

7 WDR 2003, pages 193 – 196

poverty levels and increasing general incomes, the burden on land and resources will lighten.

Dual economy

Developing countries focusing on industrialisation and/or export-led growth will undergo structural changes in moving from primary to secondary and tertiary production. This is often referred to as a transformation from a 'traditional' sector to a 'modern' sector. Developing countries will have a large rural economy based on subsistence agriculture and a growing urban sector with secondary and tertiary industries, i.e. a **dual economy**. This creates large income disparities between urban and rural areas – resulting in the mass migration from country to town seen now for decades in developing countries.

Income inequality

As covered in Chapters 83 and 86, while the world Gini coefficient has remained relatively stable over some 30 years, there has been a marked increase in income disparities between groups within many developing countries which have experienced economic growth. Figure 95.1 shows that for the largest and most populated developing countries, income distribution has, in most cases, worsened.

Summary & revision

1. Many developing countries shifted towards **market-based economies** during the 1980s.

2. Possible **benefits of market-based policies** include:

   a. Macro *economic stability* – often as the result of SAPs insisted upon by the IMF and World Bank

   b. *Increased output*, competition and income

   c. Increased *FDI and trade*

3. **Negative effects** of market-based policies have been:

   a. *Environmental degradation* and sustainability issues

   b. *Dual economies* and large urban migration

   c. Rising *income inequalities*

Figure 95.1 Change in income inequality over 20years, selected countries

(Source: OECD, "Special focus: Inequality in Emerging Economies", 2011, page 51)

DEVELOPMENT

96. Governments and Intervention

Key concepts:

- Outline of interventionist policies
- Limitations of interventionist policies
- Development and good governance
- Consensus in the end?

."…well, there's a difference of opinion on that….and if Matt were here, you'd get two more very *loud* opinions." An unnamed colleague's comment on being asked how the economics department should use additional funds.

The argument for government intervention/steering in developing countries is rather compelling; since the aim is increased prosperity and living standards, government can and should intervene to limit the negative externalities such as environmental degradation and income inequality. Note that such policies in no way negate the possibility of market systems and openness to trade. Interventionist policies should be seen as complements to market-based economic growth rather than substitutes.

Outline of interventionist policies

In summarising some of the key points given in the chapters on development, the following six points emerge as key interventionist policies:

1. Diverting government funds towards social goods, i.e. *education* and *health care*. The social and economic benefits of higher human capital have been shown in East Asia.
2. Getting the basics in place in terms of *infrastructure*.
3. Creating *solid institutions* such as banks, tax system and a fair judicial system.
4. Establishing the *rule of law* and democracy.
5. Balancing the *government budget* – which necessitates good government and a functioning tax system.
6. Putting a basic *welfare system* in place, e.g. pension schemes and basic social welfare benefit schemes.

Yes, agreed, it looks like a wish list from Neverland – and the problem is still "…how do you get in the game?" Taxes are needed for government programs and a good tax system is needed to collect taxes…and income is needed to create tax bases…and…well, you know the rest.

Limitations of interventionist policies

As looked at in Chapter 95 and several chapters in Section 3 on macro, government intervention gets mixed reviews. Many of the problems associated with government intervention are exacerbated in developing countries – primarily due to weak governments that all too frequently have been non-democratic, unaccountable and based on the power of a gun. Such governments do not establish a solid base for the six points iterated above. Instead, developing countries with *high levels of interventionism* have often seen:

- Excessive bureaucracy and red tape resulting in institutionalised corruption and inefficiency.
- Strongly interventionist policies often led to a focus on protectionist policies and infant industries where there was no comparative advantage.
- State-run monopolies which have virtually always been hugely inefficient and padded with unnecessary labour to keep unemployment figures down.
- Large scale projects tended to look more at the 'marketing side' of development rather than the extent to which the projects would increase living standards and/or income. There was a pronounced propensity of developing country governments to want to show something visible and tangible for aid funds and loans. The answer was frequently a dam, bridge or airport which was not only inappropriate for the level of development, it squandered scarce funds that could have been better used in social projects.

Development and good governance

Good governance can be defined as a system of authority giving citizens the opportunity to exercise their economic, political and legal rights. The concept encompasses a transparent, accountable and elected government body which upholds the rights of citizens and a societal consensus on how the society should look. The UN identifies eight core characteristics of good government.

1. *Participation*; men and women have the right to vote, join unions and build political parties.

DEVELOPMENT

2. *Rule of law*; fair and impartial rules laid down in a code of law and enforced impartially.

3. *Transparency*; information is open and freely available to those affected by government decisions.

4. *Consensus orientated*; government takes into account all opinions and can act as mediator.

5. *Equity and inclusiveness*; all members of society feel that they have fair shares and are regarded as equals in the eyes of government.

6. *Effectiveness*; government institutions produce the results agreed upon by the people.

7. *Accountability*; transparency and the rule of law mean that all government and private institutions can be held accountable.

Agreed, it is quite a list – and the UN admits that it is a high-order aspiration rather than a British School tick-list. Yet over a decade of evidence points to a few simple facts; good governance leads to growth and development leads to good governance. Functionally strong, open and fair governments will garner support from the populace and will not be prone to risk the ire of the very people that can vote them out of office and/or hold them accountable in courts of law. Avoiding a very tricky issue of causality here, I refer to Figure 96.1 which shows the strong correlation between good governance and income.
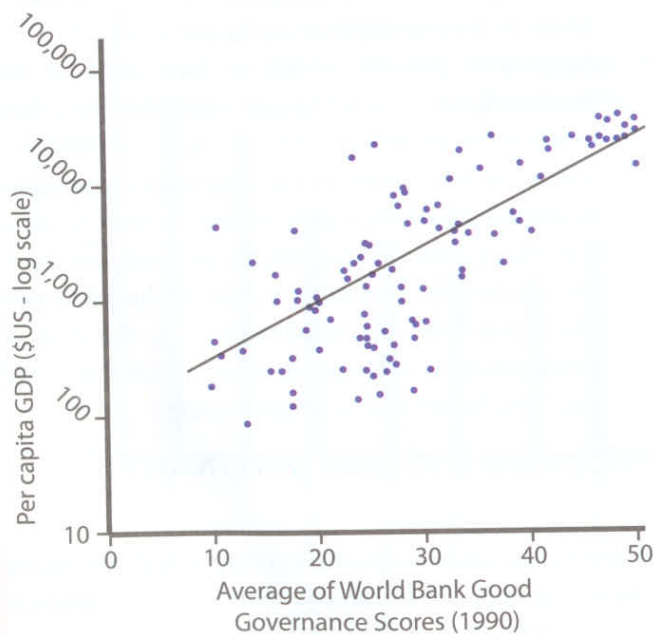
Figure 97.1 Correlation between Good Governance and Per Capita Income. (Source: World Bank Indicators 2001)

Consensus in the end?

The question remains *why* some countries have succeeded and others have failed. This is of course the 24 carat question and is at the heart of the debate concerning interventionist versus market liberalisation. There is some consensus, according to notable development economist Michael Todaro, in that successful market-orientation does not miraculously pop up in a vacuum.[1] A market system is not 'a lack of governance and rules' but quite the opposite. For a market economy to function, a number of requirements need to be met, notably along the institutional/infrastructural lines outlined earlier, e.g. financial system, legal institutions, rule of law, communications networks etc. There is also strong correlation between investment in education and the effectiveness of free market orientation. Once again it seems to come down to having a basic institutional/socio-economic framework in place before growth and development models based on market-orientation can take permanent root.

Developed countries can increase aid and direct it more efficiently by focusing on public goods and infrastructure. Wealthy countries could increase debt reduction for the countries most heavily in debt – which are also, for the most part, the poorest. Removing barriers to trade on agricultural and textile products would be enormously beneficial to developing countries; over USD20 billion a year in increased income. Finally, developed countries can increase the transfer of technology to developing countries. This includes improvements in grain varieties and farming methods and new medicines/ vaccines.

Joint responsibilities: Basically there needs to be broad-based agreement on a number of *institutions* and *rules* which govern developed and developing countries alike. Three main areas need to be addressed:

1. **Trade** rules need to be carefully outlined, and the inequity of developed country subsidies must be solved;

2. Environmental agreements must be reached which deal with the **global** problem of **negative externalities**;

3. Rules governing the protection of **international property rights** in areas such as medicines, genetically engineered crops and technology.

The 'sustainable path' is one which includes poor people and developing countries since it is simply not feasible for one fifth of the world's population to go it alone. Returning to my 'hard-line' view of development, rich countries are not 'helping the

1 See *Todaro*, page 696 ff

DEVELOPMENT

poor' but actually helping themselves – this is in essence the self-interest argument again. Developing countries are in this respect not a *subject* to render aid to but a *partner* with which to cooperate.

I finish with my favourite question, which for some reason always confuses my kids initially. "Do you want your neighbours better off or worse off?" Easy, you want them *better* off! Wealthy neighbours will take care of the sidewalks and be good to your kids. They will want to buy lemonade from your 5 year old when she opens 'Emma's Kiosk' on the clean sidewalks. They will take care of their houses and make them nice places to go for a neighbourly visit. They will watch out for burglars and other nasties. They will contribute to projects designed to make the neighbourhood a nicer, better place – if for no other reason but to increase the value of their houses.

Wealthy neighbours will be able to help *you* when you need it.

Matt
May 2012
Shady Oaks Retirement Home
Jakarta, Indonesia
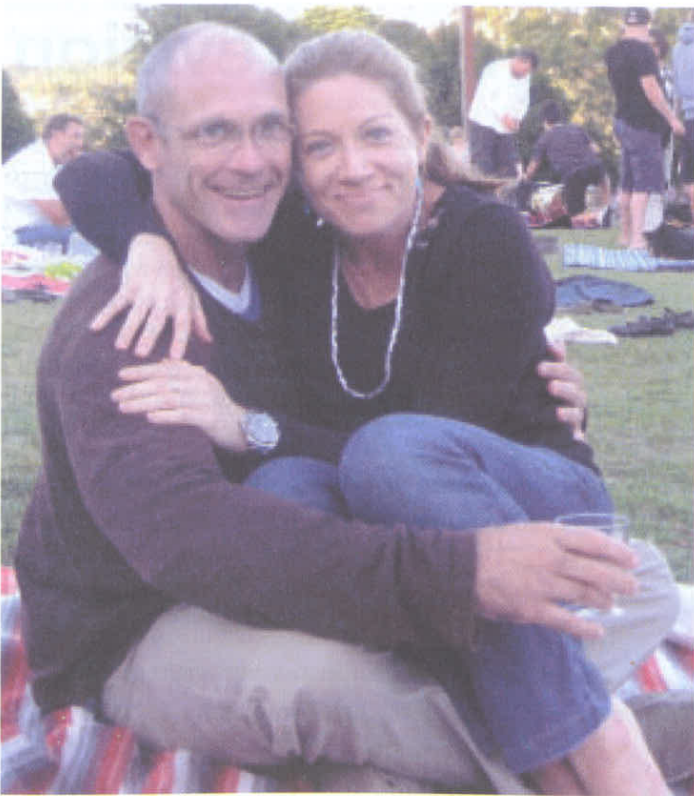
Author (left) and Bell the Very Small Australian Person. New Year's Eve 2011 on Goat Island, next to the Harbour Bridge, Sydney.

SHORT ANSWER QUESTIONS (10 MARKS EACH)

1. Why is human capital so important in the development process?

2. Why might economic growth not be compatible with sustainable development?

3. What are the possible negative consequences of economic growth in a developing country?

4. Distinguish between the various forms of aid received by developing countries.

5. Evaluate the role of multinational companies in helping developing countries to achieve economic growth/ development.

6. Explain how many developing countries became burdened with severe debt problems in the 1980s.

7. How have falling commodity prices affected many developing countries?

8. Discuss whether economic growth leads to an improvement of the environment.

9. Why are so many developing countries 'dual sector' economies?

10. Describe the 'poverty cycle' and suggest how a developing country can break the cycle.

11. Explain the importance of well-established property rights in the process of development.

12. Distinguish between 'growth' and 'development'.

13. Describe stabilisation policies as outlined by the International Monetary Fund (IMF).

14. Outline the possible negative effects of import-substitution policies.

15. Discuss whether the success of newly industrialised countries – such as the 'Asian Tiger economies' – is solely the result of outward oriented policies.

DEVELOPMENT

EXTENDED RESPONSE QUESTIONS (25 MARKS EACH)

1. (a) How might one measure differences in living standards between less developed and developed countries? (10 marks)

    (b) How might developed countries assist less developed countries to increase their living standards? (15 marks)

2. (a) Outline the main features of outward-oriented and inward-oriented development strategies. (10 marks)

    (b) Explain which of the two strategies is most likely to lead to development. (15 marks)

3. (a) What are the various forms of aid a developing country might receive? (10 marks)

    (b) Compare aid and trade as the most effective means of development. (15 marks)

4. (a) Explain how developing countries can acquire investment funds for development. (10 marks)

    (b) Evaluate the role of multinational companies (multinational enterprises) in investment and development. (15 marks)

5. (a) Explain three major barriers to development experienced by developing countries. (10 marks)

    (b) What strategies might developing countries adopt in order to overcome these barriers? (15 marks)

6. (a) Explain how a developing country can have growth without development. (10 marks)

    (b) Evaluate import-substitutions as a method of achieving growth and development. (15 marks)

7. (a) Outline the benefits of increased openness in trade. (10 marks)

    (b) Critically assess the claim that developing countries will benefit from increased trade liberalisation. (15 marks)

8. (a) Explain how foreign aid might help in the development process of a developing country. (10 marks)

    (b) Critically examine the claim that "Trade is better than aid" in the development process. (15 marks)

9. (a) Distinguish between interventionist and market-led strategies of development. (10 marks)

    (b) Evaluate the effectiveness of each approach. (15 marks)

10. (a) Examine the role of foreign direct investment (FDI) for developing countries. (10 marks)

    (b) Evaluate whether foreign direct investment is more important than aid in attaining growth/ development. (15 marks)

Summary & revision

1. Interventionist policies involve government provision of education, health care and infrastructure. A wider scope of interventionism means state-run ownership and planning of the system of production.

2. Highly interventionist governments often saw failures such as bureaucracy, trade barriers, corruption and inefficient government run monopolies.

3. Good governance has been shown to be positively correlated to economic growth.

4. Consensus has evolved in development economics, in that aid, trade and FDI can all contribute to development – given that there are functioning institutions and a degree of political stability to build on.

Index

Index

ECONOMICS
IN TERMS OF
THE GOOD, THE BAD
AND THE ECONOMIST

This is the third edition of Matt McGee's bestselling text for the I.B. Diploma course (SL & HL).

The many improved features of this edition include:

- alignment with the IB syllabus, divided into 97 'bite sized' chapters.
- where appropriate, large colour coded diagrams linking text and theory.
- revision sections, including core diagrams, at the end of chapters.
- substantial additional material. eg exam-style questions & answers, case studies, ... (IA and EE will be available as downloads from the IBID website).
- non-syllabus material clearly indicated as 'A Little Depth' for students who want to take the subject further.

Students and teachers who are used to a 'dry' treatment of Economics will find this book provocative, enlightening and refreshing.

This is a rather personal account from an economist who clearly loves teaching and is equipped with a wry sense of humour.

Readers will find a mixture of solid academia, shoot-from-the-hip earthiness and simple common sense.

ISBN: 978-1-921917-02-8