

THE NATURE AND SCOPE OF ECONOMICS.....	3
The scope of economics.....	3
The tools of economic analysis.....	5
Alternative economic systems.....	8
Consumer behaviour.....	9
PRICE THEORY, SUPPLY AND DEMAND.....	12
Markets and competition.....	12
Demand.....	14
Supply.....	21
The price mechanism.....	25
Elasticity Of Demand And Supply.....	27
Price elasticity of demand.....	27
Other elasticities of demand.....	32
Elasticity of supply.....	33
Market Failure: Externalities And Intervention.....	36
Market imperfections and market failure.....	36
Externalities and public goods.....	37
Price regulation.....	39
Indirect taxes and subsidies.....	42
Theory Of The Firm: Production And Costs.....	46
Costs of production and the firm.....	46
Average costs, marginal costs and diminishing returns.....	49
Profit maximisation.....	51
Economies of scale and long-run costs.....	54
Factor Markets And Factor Rewards.....	58
Factors of production and their rewards.....	58
Capital and interest.....	59
Labour and wages.....	60
Land and rent.....	70
Entrepreneurship and profit.....	72
Market Structures: From Perfect Competition To Monopoly.....	74
Equilibrium under perfect competition.....	75
Equilibrium for a monopoly.....	77
Monopoly and perfect competition compared.....	80
Price discrimination.....	82
Public policy towards monopolies.....	83
Privatisation of nationalised industries.....	86
Monopolistic Competition And Oligopoly.....	87
Monopolistic competition and non-price competition.....	87
Oligopoly.....	90
Money and Banking.....	94
Monetary aggregates.....	97
Flows of funds within an economy.....	98
Banking.....	101
The building societies.....	106
The Central Bank And The Financial Markets.....	107
The functions of the Bank of England.....	107
The Bank of England as a banker.....	108
The capital markets and the money markets.....	110
NATIONAL INCOME ANALYSIS.....	115
Measuring the creation of economic wealth.....	115
The circular flow of income.....	116
National income statistics.....	118
NATIONAL INCOME AND UNEMPLOYMENT.....	123
Aggregate demand and supply and equilibrium national income.....	123
Consumption, savings and investment.....	126
Unemployment And Inflation.....	131
Unemployment.....	131
The Phillips curve.....	132

<u>Inflation and its consequences.....</u>	<u>135</u>
<u>Prices and incomes policy.....</u>	<u>137</u>
<u>The Demand For And Supply of Money.....</u>	<u>139</u>
<u> The quantity theory of money.....</u>	<u>139</u>
<u> Keynesian views on the demand for money.....</u>	<u>143</u>
<u> The monetarist viewpoint.....</u>	<u>145</u>
<u> A comparison of the two views.....</u>	<u>148</u>
<u>Fiscal Policy.....</u>	<u>150</u>
<u> Fiscal policy, national income and demand management.....</u>	<u>150</u>
<u> Taxation policy and its effects.....</u>	<u>154</u>
<u>International Trade.....</u>	<u>159</u>
<u> Absolute and comparative advantage.....</u>	<u>159</u>
<u> Free trade and its advantages.....</u>	<u>161</u>
<u> Free trade and protection.....</u>	<u>162</u>
<u>The Balance Of Payments. Exchange Rates.....</u>	<u>168</u>
<u> The balance of payments.....</u>	<u>168</u>
<u> Components of the balance of payments.....</u>	<u>169</u>
<u> The Terms of Trade.....</u>	<u>172</u>
<u> Exchange rates.....</u>	<u>173</u>
<u> Exchange rate policies of governments.....</u>	<u>176</u>
<u> European monetary cooperation.....</u>	<u>177</u>

www.allonlinefree.com

THE NATURE AND SCOPE OF ECONOMICS

Your objectives

After completing this chapter you should be aware of

- (a) the problems which the social science of economics attempts to address;
- (b) the basic tools of economic analysis;
- (c) the alternative systems which have been devised to tackle economic problems;
- (d) the assumptions about consumer behaviour which shape the solutions to economic problems.

The scope of economics

A social science deals with some aspect or aspects of human society. Economics is a social science which is concerned with the allocation of scarce resources to provide goods and services which meet the needs and wants of consumers.

Rationality

One of the important assumptions in economics, and one on which much economic theory is based, is the rationality of human behaviour. In order to make predictions about their economic behaviour, economists assume that human behaviour is 'rational' and that consumers and producers act rationally. For example, producers and consumers will make reasoned decisions about how much to produce or buy at any given price.

The optimum

The assumption of the rationality of human behaviour, and that people will take decisions and actions which are directed towards a rational objective, leads us to the concept in economics of the optimum. The optimum means the best possible, and the following are underlying assumptions in much economic analysis.

- (a) Producers will seek to maximise their profits and returns.
- (b) Consumers will seek to maximise the benefits they obtain (their 'utility') from using the income at their disposal.
- (c) Governments will seek to maximise the well being of their population (for example, by maximising the national income per head of the population).

Positive and normative economics

You might already have strong personal views about what sort of economic society we should have, e.g. whether a free market 'capitalist' economy is desirable, or whether a centrally planned 'command' economy is preferable. In the study of economics, it is easy for us to be influenced in our views by our ideas of 'what ought to be'.

Economists also have views on these subjects, and some economic writing is aimed at influencing decision makers by prescribing actions, which, in the opinion of the author, will lead to ends, which he or she considers desirable. Other economic research, however, is directed purely to finding out what the consequences will be if certain actions are taken, without expressing any view on the desirability of those consequences. These two different approaches are referred to as normative economics and positive economics.

- (a) Normative economics is concerned with the expression of value judgements by economists, of what they would like to happen -e.g. what sort of economic society they would like to see in operation.
- (b) Positive economics is concerned with objective statements about what does happen or what will happen. A positive approach is more objective, and more scientific, and it is the approach we shall try to take in our study of economics here.

Microeconomics and macroeconomics

The study of economics is divided into two halves, microeconomics and macroeconomics.

- (a) 'Micro' comes from the Greek word meaning small, and microeconomics is the study of individual economic units or particular parts of the economy -e.g. how does an individual household decide to spend its income? How does an individual firm decide what volume of output to produce or what products to make? How is the price of an individual product determined? How are wage levels determined in a particular industry?
- (b) 'Macro' comes from the Greek word meaning large, and macroeconomics is the study of 'global' or collective decisions by individual households or producers. It looks at a national or international economy as a whole -e.g. total output, income and expenditure, unemployment, inflation, interest rates and the balance of international trade etc, and what economic policies a government can pursue to influence the condition of the national economy.

In this text, we shall study microeconomics first before going on to look at the macroeconomic environment in later chapters.

The fundamental problem of economics

A fundamental concept in economics is the scarcity of resources. There are not enough resources to meet the needs of consumers and producers.

- (a) In the case of consumers, the scarcity of goods and services might seem obvious enough. Everyone would like to have more -another car, a bigger home, more domestic goods, better food and drink, more holidays, more trips to the cinema or theatre, a boat, a private plane, membership of more clubs and societies, more clothes and so on. There simply isn't enough to go round to satisfy the potential demand.
- (b) In the case of producers, there are four scarce resources:
 - (i) natural resources, referred to collectively as 'land';
 - (ii) labour;
 - (iii) capital -e.g. equipment and tools;
 - (iv) enterprise or entrepreneurship.

Scarce resources mean that producers cannot make unlimited Quantities of goods and services. Since resources for production are scarce and there are not enough goods and services to satisfy the total potential demand, choices must be made. Choice is only necessary because resources are scarce.

- (a) Consumers must choose what goods and services they will have.
- (b) Producers must choose how to use their available resources, what goods and services to produce, and in what quantities.

The fundamental problem of economics is the allocation of these scarce resources. What will be produced? What will be consumed? And who will benefit from the consumption?

The tools of economic analysis

In going about their work economists have to deal with a number of variables. To the economist, a variable is anything that influences the basic decisions, which are the subject of his science: what will be produced, what will be consumed and who will benefit from the consumption? Economic variables therefore include, amongst many others:

- (a) prices of commodities and of services;
- (b) quantities of production resources available;
- (c) numbers of skilled and unskilled workers;
- (d) consumer attitudes (since these can affect decisions about spending and saving),

To assist them in their analysis of variables, economists need to accumulate data about them: price levels, stock counts, employment statistics, consumer surveys and so on.

Often it will be convenient to represent the variables under discussion by means of a model. At its simplest, this process may mean no more than drawing a graph and indeed you will find that the use of graphs is a very common technique in economics.

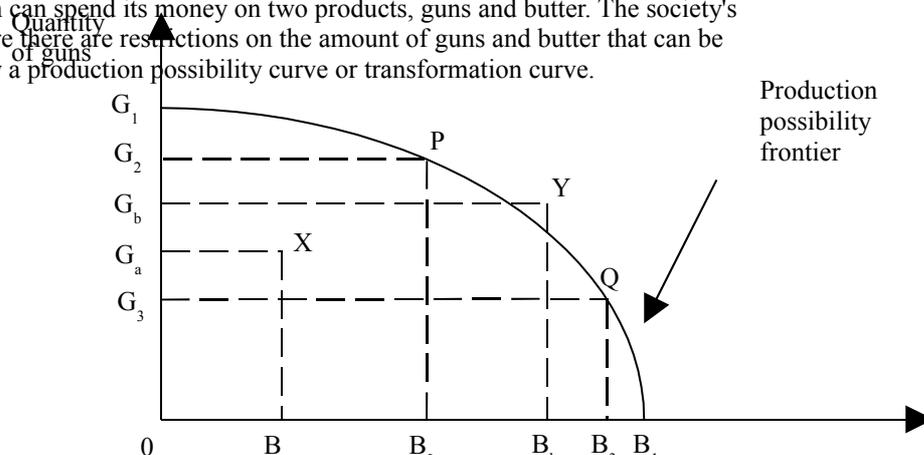
But economists also use much more complex models than this. Later in this book, for example, we develop a model, which attempts to show the way in which income circulates among the households and firms in an economy.

One point which may already have struck you is that no matter how complex the model becomes, no matter how many variables the economist tries to incorporate, the end result will never be more than an approximation of the real world. This is true even of the complicated computerised models used by government economists in their attempts to predict economic trends. The complexity of a real-life economy is simply too great to be simulated in a model.

Despite this, economic models (even the simplest graphs and equations) are of great value. The very complexity of a real economy is a barrier to understanding. By using models, economists are able to isolate, from the great mass of raw economic activity, the variables that appear relevant to any particular problem. Models enable the economist to understand and analyse a situation and, eventually, to make predictions about future economic events.

The production possibility curve

We will begin our examination of the fundamental problem of economics by using a very simple model: that of a society which can spend its money on two products, guns and butter. The society's resources are limited: therefore there are restrictions on the amount of guns and butter that can be made, which can be shown by a production possibility curve or transformation curve.



The curve from G, round to B, shows the various combinations of guns and butter that a society can make, if it uses its limited resources efficiently.

- (a) The firm can choose to make up to:
- (i) G_1 units of guns and no butter;
 - (ii) B_1 units of butter and no guns;
 - (iii) G_2 units of guns and B_2 of butter (point p on the curve);
 - (iv) G_3 units of guns and B_3 of butter (point Q on the curve).

Points P and Q on the curve are chosen at random. Any other point on the curve would indicate production of another possible combination of guns and butter .

- (b) The combination of G_a units of guns and H_a units of butter plotted at point X is within the production possibility curve. More than these Quantities can be made of either or both guns and butter. Point X is therefore an inefficient production point for the economy, and if the society were to make only G_a of guns and H_a of butter, it would be under-utilising its resources.

The production possibility curve illustrates the need to make a choice about what to produce (or buy) when it is not possible to have everything -i.e. when there is scarcity. Although we have characterised the products of our hypothetical economy as guns and butter, we can generalise the production possibility curve to show the production possibilities for different types of good, as for 'good X' on one axis and 'all other goods' on the other axis.

Choice involves sacrifice. If there is a choice between having guns and having butter, and a country chooses to have guns, it will be giving up butter to have the guns. The cost of having guns can therefore be regarded as the sacrifice of not being able to have butter.

There is a sacrifice involved in the choices of consumers and producers, as well as the choices of governments.

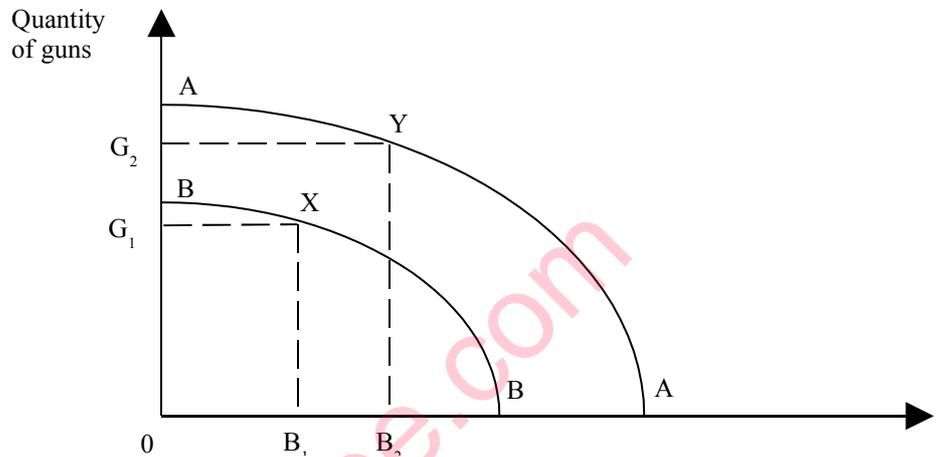
Suppose a consumer has a limited amount of money, and chooses to buy some eggs. One measurement of the cost of the eggs is their money price. Another way of looking at the cost is to consider the sacrifice involved in choosing eggs rather than, say, milk. If the consumer has some eggs, he or she is giving up the opportunity to have some milk, and the benefits that the milk would have provided. The cost of an item measured in terms of the alternatives forgone is called its opportunity cost. Thus the opportunity cost of buying six eggs can be measured as the two pints of milk or the one bus ride that could have been bought instead. Similarly, the opportunity cost of a country having a nuclear missile could be measured in terms of the number of schools that could have been built and staffed with the same amount of resources

At the level of the firm, the production possibility curve can be seen as showing the maximum output of different alternative goods which the firm can produce when all of its resources are fully used -for example, a firm might operate production lines capable of producing washing machines or refrigerators. Producing more washing machines bears the opportunity cost of a lower level of production of refrigerators.

Shifts in the production possibility curve

When the availability of resources changes, or there is a development in technology, the production possibility curve may shift. Changes are made possible by developments such as a bigger labour force, more efficient methods of working, more efficient machinery, or a new discovery of natural resources, such as oil, natural gas or minerals.

- (a) If the production possibility curve moves outwards, to the right, it means that the economy is capable of producing more goods and services in total than it could before, and there is economic growth.
- (b) If it moves to the left (inwards) it means that the economy cannot produce as much as before (e.g. because of a significant decline in population or the exhaustion of a natural resource).



In Figure 2, curve AA represents greater production possibilities than curve BB. If a society's production possibility curve shifts out from BB to AA, there is economic growth. The society could now switch from making G_1 of guns and B_1 of butter (point X) to making G_2 of guns and B_2 of butter (point Y).

Exchange value

When a resource is scarce (that is, when less of it is available than is needed to satisfy the wants and needs of producers and consumers), it has an economic value or an 'exchange' value. Producers will give something in exchange for the natural resources, labour and equipment that they need to help them to produce goods and services. Consumers will give something in exchange for the goods and services that they can obtain.

In an advanced economy, value is measured in money terms -e.g. the price of goods, wage levels, and the cost of raw materials. But similar principles apply in a primitive barter economy. (A barter economy is one in which goods are traded for other goods, without the use of an exchange medium such as money.) For example, if a potato grower were to ask someone to help him to dig up the potatoes he had grown, he might offer some of the potatoes to the helper in exchange for the labour the helper puts in.

For resources to have an exchange value, there is a presupposition that exchange can take place. In a market economy, this is what does happen.

- (a) Resources are owned, sometimes by society as a whole, and often by individuals, or groups, or organisations. Individuals, for example, own their own labour. Property owners own land, and the minerals in them.
- (b) Owners of resources will give some of their resources in exchange for others. Individuals, for example, will give their labour in exchange for a wage. In a barter economy, a farmer might give up some land in exchange for some horses or cows.
- (c) In an advanced market economy, an important medium of exchange is notes and coin.

The division of labour

Significant features of all 'modern' economies are specialisation and the division of labour, both of which increase the need for exchange.

- (a) Specialisation of labour occurs where a worker makes just one type of product or provides one type of service -e.g. a farmer, a doctor, an accountant, a butcher, an entertainer, and so on. Individuals are not self-sufficient and do not each make themselves all the products and services they require for their personal wants. Instead, individuals specialise in making a particular good or service.
- (b) The division of labour refers to specialisation within a single industry. The manufacture of one product sometimes calls for the work to be divided up into 'sub-specialities'; for example, in the making of motor cars, there are car body builders, paint sprayers and engine tuners.

The existence of specialisation and division of labour necessitates exchange in a modern market economy, because specialists are not self-sufficient, and they must exchange what they produce in return for other goods and services they need. In practice, most workers sell their labour to a firm (or to the government) in exchange for money wages. Money is the medium of exchange, which permits them to buy other goods and services.

Alternative economic systems

Means of allocating resources

Scarcity of resources means that choices must be made about how the resources will be allocated. There are three basic resource allocation decisions.

- (a) What goods and services should be produced? This will depend on what consumers want to buy, and what they will pay for each product or service. The decisions about what will be produced relate to demand and supply:
 - (i) demand means the demand from customers or consumers, and satisfied demand is actual consumption;
 - (ii) supply of goods and services is referred to as production.
- (b) How will these goods and services be produced? The producers or suppliers of goods and services might be small companies, large companies, monopolies, state-owned enterprises or the government itself.

The choice about who will produce the goods and services, and what mix of resources the producers will use, will depend on the costs of resources and the efficiencies of resource utilisation.
- (c) To whom will the goods and services be distributed? Some goods and services are provided free by the state (e.g. in the UK, some health care and education) but others have to be paid for. The distribution of goods and services will therefore depend on the distribution of income and wealth in society. This in turn will depend on what individuals and organisations earn, and the theory of distribution in economics is concerned with what rewards are earned by the owners of scarce economic resources: land, labour, capital and entrepreneurship.

The way in which these resource decisions are resolved depends on the type of economy we are dealing with. In this context, a number of terms are explained briefly below and will be developed more fully in later chapters.

- (a) In a free market economy, the decisions and choices about resource allocation are left to market forces of supply and demand, and the workings of the price mechanism. In other words, what producers will make and what consumers will buy are kept in balance by the price that producers will want for their output and the price that consumers are willing to pay.

- (b) In a centrally planned economy or command economy, the decisions and choices about resource allocation are made by the government. Money values are attached to resources and to goods and services, but it is the government that decides what resources should be used, how much should be paid for them, what goods should be made and what their price should be.
- (c) In a mixed economy the decisions and choices are made partly by free market forces of supply and demand, and partly by government decisions. All national economies are mixed economies, although with differing proportions of free market and centrally planned decision-making from one country to the next.

Consumer behaviour

In this section we look at how consumers make economic choices, so that we can then go on to examine how those choices interact together in economic markets.

A key concept in the study of consumer behaviour is utility. Utility is the word used to describe the pleasure or satisfaction or benefit derived by a person from the consumption of goods. Total utility is then the total satisfaction that a person derives from spending his income and consuming goods.

Marginal utility is the satisfaction gained from consuming one additional unit of a good or the satisfaction forgone by consuming one unit less. If someone eats six apples and then eats a seventh, total utility refers to the satisfaction he derives from all seven apples together, while marginal utility refers to the additional satisfaction from eating the seventh apple, having already eaten six.

Assumptions about rationality

We need to make some additional assumptions about the rational behaviour of the consumer, which was discussed earlier:

- (a) generally the consumer prefers more goods to less;
- (b) generally the consumer is willing to substitute one good for another provided its price is right;
- (c) choices are transitive. This means that if at a given time a commodity bundle A is preferred to bundle B and bundle B is preferred to bundle C then we can conclude that commodity bundle A is preferred to commodity bundle C.

Acting rationally means that the consumer attempts to maximise the total utility attainable with a limited income. When the consumer considers whether any unit of a good is worth buying he is deciding whether the marginal utility of buying another unit of the good exceeds the marginal utility that would be yielded by any alternative use of the same amount of money.

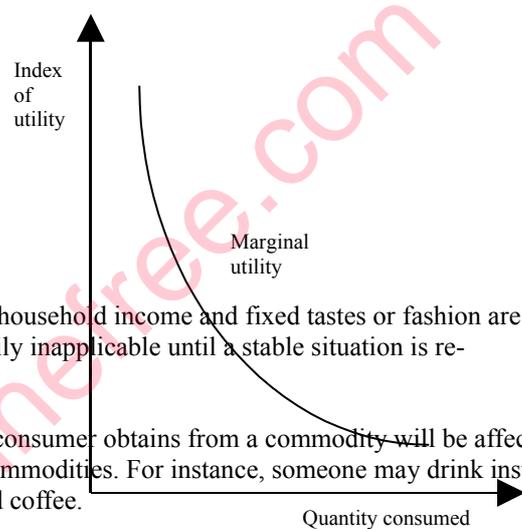
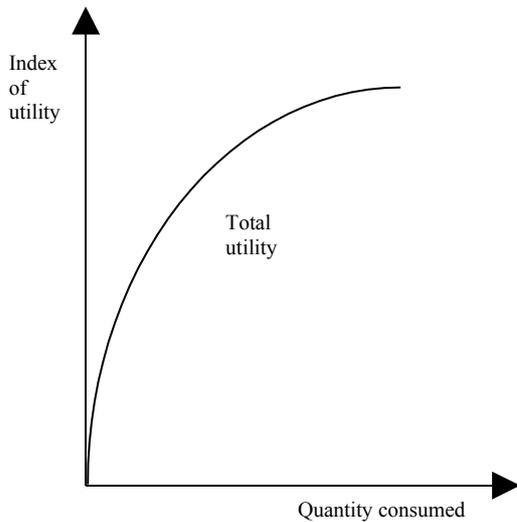
If a person has maximised his total utility, it follows that he has allocated his expenditure in such a way that the utility gained from spending the last penny on each of the commodities he buys will be equal, i.e. the consumer will spend his income in such a way that he gets the same marginal utility from the last penny spent on each commodity.

Diminishing marginal utility

As a person consumes more of a commodity, the total satisfaction he gains will continue to increase, but the marginal utility derived from increasing consumption will fall with each additional unit consumed. The earlier example of a man on holiday provides an illustration of this. The total satisfaction he gains will increase as his holiday gets longer. He is unlikely, however, to derive as much utility from the second day as much as the first, or the third as much as the second, and so on.

The law of diminishing marginal utility states that, all other things being equal, the additional satisfaction derived from consuming additional units of a commodity will diminish with each

successive unit consumed. Total utility will continue to rise as each successive unit is consumed, but at a decreasing pace.



The 'law' applies only if the assumptions of fixed household income and fixed tastes or fashion are valid. If either changes, the 'law' will be temporarily inapplicable until a stable situation is re-established.

- (a) Income: if income changes, the utility a consumer obtains from a commodity will be affected by the changing consumption of other commodities. For instance, someone may drink instant coffee and switch to buying fresh-ground coffee.
- (b) Tastes: a change in tastes, fashion or attitudes may occur as more of a commodity is consumed and marginal utility may increase where such a change is taking place. For example, a person may progress from an occasional buyer of pictures into an obsessive art collector.

Consumer equilibrium, marginal utilities and relative prices

We stated earlier that a consumer will maximise his total utility, with a given income and tastes, at a level of consumption where the marginal utility from the last penny spent is the same for each commodity bought.

This proposition can be developed into an algebraic formula. Suppose that a household buys two commodities, X and Y.

Let the marginal utility of a unit of X be MU_x
 and the marginal utility of a unit of Y be MU_y
 Let the price per unit of X (in pence) be P_x
 and the price per unit of Y be P_y

The household will attain a utility-maximising equilibrium where the marginal utility from the last penny spent is the same for X and Y, i.e. where

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} \dots\dots\dots(1)$$

Cross multiplying gives: $\frac{MU_x}{MU_y} = \frac{P_x}{P_y} \dots\dots\dots(2)$

This is true for any pair of commodities bought by the household.

The law of equi-marginal returns

In order to maximise their total utility, people will distribute their expenditure in such a way that the ratio of marginal utilities for all the goods they consume ($MU_x: MU_y: MU_z$ etc) of goods is equal to the relative price of the goods (i.e. the ratio of prices $P_x: P_y: P_z$). This is sometimes called the law of equi-marginal returns. It is a statement of the principle that a utility-maximising household will allocate its expenditure so that the marginal utility of that last penny spent on each good is equal for each good that it buys. This is how the household decides what quantities of each good it would want to buy, given the existing prices of all goods.

We can use the concept of equi-marginal returns and the equilibrium relationship expressed in equation (2) above to predict what will happen to consumer spending patterns when there is a change in the price of a commodity.

Price rises and changes in income

In speaking of a price rise, we mean a rise in the real price, as distinct from the money price or nominal price of a commodity. If all prices rise in the same proportion because of inflation, there would be a rise in the nominal prices, but real prices would remain the same because relative values would be unaltered. If the price of a good has risen more than those of other goods and incomes generally, then a real price rise has occurred.

We saw above that the household is in equilibrium when the ratio of marginal utilities of commodities is equal to the ratio of their prices. We can therefore predict that if the price of a commodity rises, the marginal utility of the commodity per penny spent will fall, i.e. if the price of a commodity X rises while the price of y remains the same, MU_x/P_x will fall. Therefore $MU_x/P_x < MU_y/P_y$ and so $MU_x/MU_y < P_x/P_y$ (note: $<$ means 'is less than'). The price rise alters the consumer's equilibrium.

This analysis leads to a basic prediction of demand theory, which we shall be examining in the next chapter.

- (a) If the price of a commodity rises (with income, tastes and all other prices constant), demand by each household for that commodity will fall.
- (b) If the price of a commodity falls (with income, tastes and all other prices constant), demand by each household for that commodity will rise.

PRICE THEORY, SUPPLY AND DEMAND

Your objectives

After completing this chapter you should:

- (a) understand that the supply of and demand for commodities is influenced by competitive factors
- (b) be aware of the factors that stimulate or depress supply of and demand for a commodity
- (c) understand how the price of a commodity is determined by the interaction of supply and demand

Markets and competition

Introduction

In microeconomics, the theory of consumer behaviour, the theory of the firm and price theory seek to explain how economic decisions are reached.

- (a) What makes consumers decide what they are going to buy, and in what quantities?
- (b) What makes firms decide what goods and services they are going to produce, and in what quantities?
- (c) How are the market prices for buying and selling arrived at?

The concept of a market

The concept of a market in economics goes beyond the idea of a single geographical place where people meet to buy and sell goods. It is a term used to refer to the buyers and sellers of a good who influence its price. Markets can be worldwide, as in the case of oil, wheat, cotton and copper, for example. Others are more localised, such as the housing market or the market for second-hand cars.

Markets for different goods or commodities are often inter-related. All commodities compete for households' income so that if more is spent in one market, there will be less to spend in other markets. Further, if markets for similar goods are separated geographically, there will be some price differential at which it will be worthwhile for the consumer to buy in the lower price market and pay shipping costs, rather than buy in a geographically nearer market.

Decision-takers in a market

There are two major groups of decision-takers in a market, namely buyers and sellers, or more accurately:

- (a) purchasers and would-be purchasers;
- (b) suppliers and would-be suppliers.

Suppliers and would-be suppliers are referred to in economics as firms. Buyers and would-be buyers are often referred to in economics as households. We often refer to 'households' rather than to individual consumers partly because economic data that is collected refers to household financial transactions rather than individual consumers' spending. We can then ignore all problems of decision-making within the household (i.e. which member of the household makes the spending decisions) and regard the household as though it were a single individual.

Some markets have buyers who are not households at all, but who are other firms or government authorities. For example, a manufacturing firm buys raw materials and components to go into the products that it makes. Service industries and government departments must similarly buy in supplies in order to do their own work.

However, the demand for goods from firms and government authorities is a derived demand in the sense that the size of their demand depends on the nature of the demand from households for the goods and services that they in turn produce and provide.

To begin with, we shall concentrate on product markets, which are markets in which a good or a service to consumers is bought and sold.

Later, we shall go on to look at resource markets, which are the markets in which production resources -especially labour and capital -are bought and sold.

Price theory

Price theory (or demand theory as it is sometimes called) is concerned with how market prices for goods are arrived at, through the interaction of demand and supply.

The economist distinguishes between perfect and imperfect competition in markets. All markets have some imperfections, but the perfect market provides a useful theoretical benchmark or starting point for assessing the characteristics of a market in the 'real world',

Perfect competition

In a perfect market for a product:

- (a) there is a large number of buyers and a large number of sellers and no individual buyer or seller can influence the market price. An individual firm must accept the prevailing market price -i.e. it must be a 'price taker';
- (b) there is perfect communication so that all buyers and sellers have the same information about prices through the market, and buyers and sellers can obtain this information without cost;
- (c) the consumer will act rationally and will therefore try to pay the lowest price at which a product is offered. The producer, acting rationally, will try to get the highest possible price for his product in order to maximise his profits;
- (d) the product is homogeneous, i.e. it is uniform across the market and there is no product differentiation so that one firm cannot sell a product similar to its competitors' products by emphasising or advertising its differences or brand image;
- (e) there is freedom of entry into the market by new sellers;
- (f) there is an absence of transport costs in travelling between one part of the market and another.

These conditions ensure that price differences within the market are rapidly eliminated and a single price is established throughout the market for all products sold in the market.

Market imperfections

Although organised markets come close to the theoretical model of a perfect market, perfect markets do not exist in the 'real world', and actual markets are all imperfect to some degree. There are the following reasons for this.

- (a) Buyers and sellers usually have incomplete information about prices ruling in all other parts of the market. For instance, a shopper may find it inconvenient to check the price of strawberries in every local shop, and might as a result buy at a price above the lowest obtainable.

- (b) Producers can create the impression that their goods are better than those of their competitors, although they are really quite similar. Such product differentiation is achieved not only by differences in product design but also by means of advertising and branding.
- (c) Customer loyalty or inertia sometimes prevents rational decisions by buyers. Customers might continue to go to a supplier who has given good service in the past rather than tryout a new and cheaper competitor.
- (d) Many markets are not perfectly competitive. Perfect and imperfect competition will be discussed in a later chapter. However, the essence of the distinction is that in an imperfect market there is an imbalance in economic power. A group of suppliers or buyers may be able to influence total market quantities supplied or demanded and the price of the product.

For our immediate purpose, however, a perfect market is assumed to exist. Although this is an over-simplification, it helps to provide useful insights into price determination. The assumption of perfect markets will be relaxed in later chapters.

We shall now look at demand and supply in turn, and then consider how demand and supply interact through the price mechanism.

Demand

The concept of demand

Demand refers to the Quantity of a good that potential purchasers would buy, or attempt to buy, if the price of the good were at a certain level.

It is important that you should appreciate the concept of demand properly. Demand does not mean the quantity that potential purchasers wish they could buy. For example, potential purchasers might desire to have one million units of a good, but there might only be actual attempts to buy one hundred units at a given price.

Demand is what would be the actual attempt to buy at a given price. Demand might be satisfied, and so actual Quantities bought would equal demand. On the other hand, some demand might be unsatisfied, with more would-be purchasers trying to buy a good that is in insufficient supply, and so there are not enough units of the good to go around.

Several factors influence the total market demand for a good. One of these factors is obviously its price, but there are other factors too, and to help you to appreciate some of these other factors, you need to recognise that households buy not just one good with their money but a whole range of goods and services.

The factors on which the quantity of demand for an individual good is dependent include:

- (a) the price of the good;
- (b) the price of other goods (products and services);
- (c) the size of household income;
- (d) tastes and fashion;
- (e) expectations;
- (f) the distribution of wealth amongst households -i.e. how wealth is spread amongst the population.

Each of these factors is discussed at greater length in the paragraphs that follow.

The total quantity demanded is referred to as the market demand. Whereas factors (a) to (e) can affect buying decisions by individual households, factor (f) important because it influences the potential aggregate size of the market as a whole.

Demand and the price of a good

In the case of most goods (some exceptions, such as Giffen goods, will be discussed later), the higher the price, the lower will be the quantity demanded, and the lower the price, the higher will be the quantity demanded. It is common sense that at a higher price, a good does not give the same value for money as it would at a lower price, and so households would not want to buy as many. This dependence of demand on price applies to all goods and services, from bread and salt to houses and space rockets.

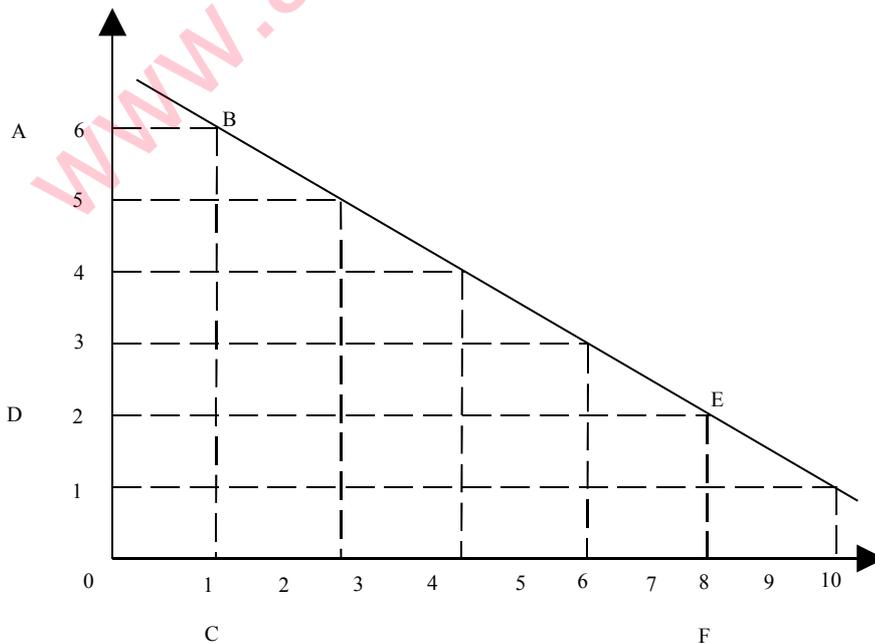
The demand curve

The relationship between demand and price can be shown graphically as a demand curve. The demand curve of a single consumer is derived by estimating how much of the good the consumer would demand at various hypothetical market prices. Suppose a consumer has the following demand schedule for soap powder.

<i>Price per kilogram</i>	<i>Quantity demanded (kilos)</i>
£ 1	9 $\frac{2}{3}$
2	8
3	6 $\frac{1}{4}$
4	4 $\frac{1}{2}$
5	2 $\frac{2}{3}$
6	1

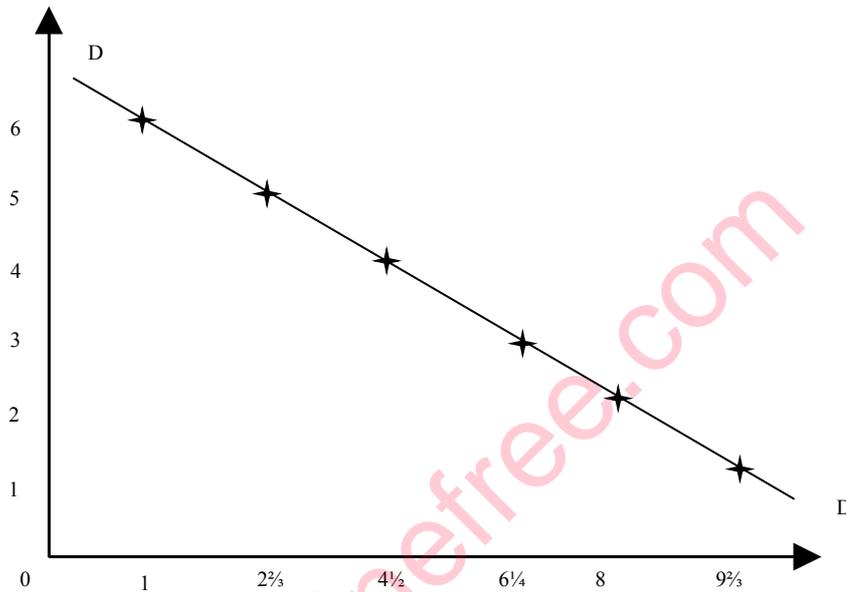
We can show this schedule graphically, with:

- (a) price on the vertical axis; and
- (b) quantity demanded on the horizontal axis.

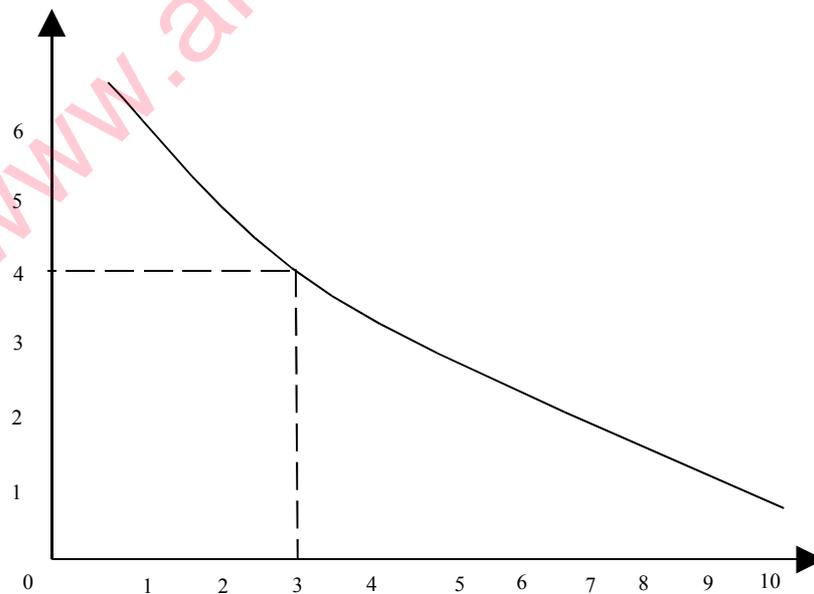


The area of each rectangle represents consumers' total money outlay at the price in question. For example, at a price of £6, demand would be 1 kilogram and total spending would be £6, represented by rectangle ABCO. Similarly, at a price of £2, demand would be 8 kilograms and the total spending of £16 is represented by rectangle DEFO.

If we assume that there is complete divisibility, so that price and quantity can both change in infinitely small steps, we can draw a demand curve by joining the points shown in Figure 1 by a continuous line, DD. This is the consumer's demand curve for soap powder in the particular market we are looking at.



Here the demand curve happens to be a straight line. Straight line demand curves are often used as an illustration in economics because it is convenient to draw them this way. In reality, a demand curve is more likely to be a curved line convex to the origin. A convex demand curve will mean that there are progressively larger increases in Quantity demanded as price falls.



Changes in demand caused by changes in price are represented by movements along the demand curve, from one point to another. The price has changed, and the quantity demanded changes, but the demand curve itself remains the same.

In this simple example, we are looking at the demand schedule of a single household. A market demand curve is a similar curve, drawn from a demand schedule, expressing the expected total Quantity of the good that would be demanded by all consumers together, at any given price.

Market demand refers to the total Quantities of a product that all households would want to buy at each price level. A market demand schedule and a market demand curve are therefore simply the sum of all the individual demand schedules and demand curves put together. Market demand curves would be similar to those in Figures 2 and 3, but with Quantities demanded being higher -i.e. total market demand.

The market demand curve generally slopes down from left to right because:

- (a) for the individual consumer, a fall in the price of the good makes it relatively cheaper compared to other goods and with his limited budget, expenditure will be shifted to the good whose price has fallen. It is the relative price of the good that is important. A fall in the relative price of a good increases demand for it. This is referred to as the substitution effect;
- (b) a fall in the good's price means that people with lower incomes will also be able to afford it. The overall size of the market for the good increases. The converse argument applies to an increase in prices; as a price goes up, consumers with lower incomes will no longer be able to afford the good, or will buy something else whose price is relatively cheaper, and the size of the market will shrink. This is referred to as the price effect.

This analysis indicates that the relationship between price and quantity demanded is an inverse one: as price rises, demand falls; as price falls, demand increases.

A demand curve shows how the quantity demanded will change in response to a change in price provided that all other conditions affecting demand are unchanged -i.e. provided that there is no change in the prices of other goods, tastes, expectations or the size of household income. (This assumption, that all other variables remain unchanged as we examine changes in the variable that interests us, is often referred to by use of the Latin phrase *ceteris paribus* -other things being equal.)

Substitutes and complements

A change in the price of one good will not necessarily change the demand for another good. For example, we would not expect an increase in the price cocoa to affect the demand for motor cars. However, there are goods for which the market demand is in some way inter- connected. These inter-related goods are referred to as either substitutes or complements.

- (a) Substitute goods are goods that are alternatives to each other, so that an increase in the demand for one is likely to cause a decrease in the demand for another. Switching demand from one good to another 'rival' good is substitution. Examples of substitute goods are:
 - (i) rival brands of the same commodity, e.g. Coca-Cola and Pepsi-Cola;
 - (ii) tea and coffee;
 - (iii) bus rides and car rides;
 - (iv) different forms of entertainment.

Substitution takes place when the price of one good rises or falls relative to a substitute good.

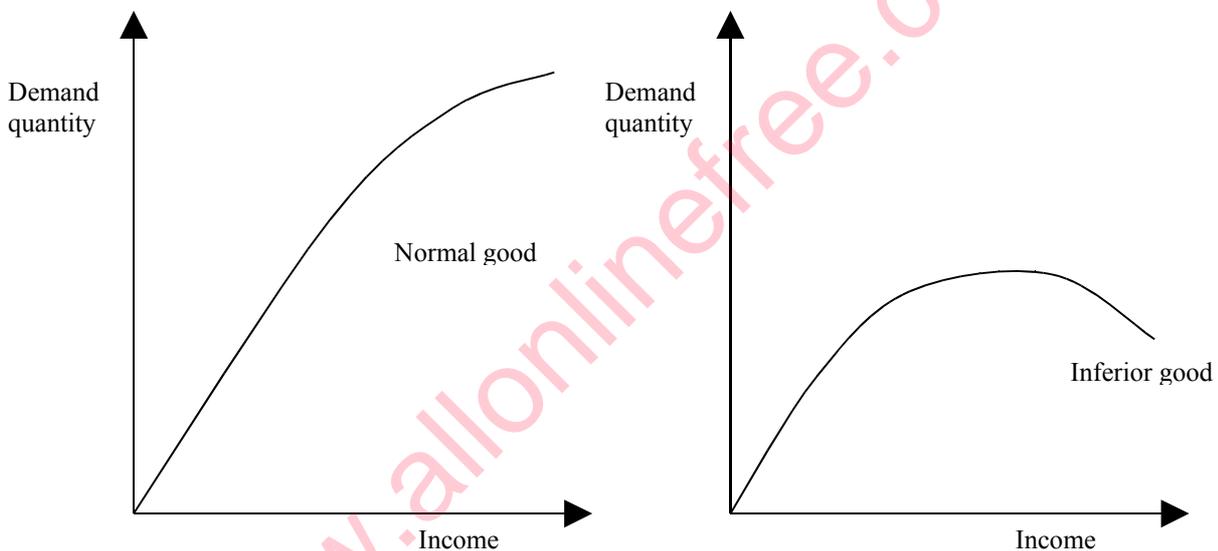
- (b) Complements are goods that tend to be bought and used together, so that an increase in the demand for one is likely to cause an increase in the demand for the other. Examples of complements are:
 - (i) cups and saucers;
 - (ii) bread and butter;
 - (iii) motor cars and motor spares.

Demand and household income

The amount of income that a household earns will affect the demand for a good. As you might imagine, more income will give households more to spend, and they will want to buy more goods at existing prices. However, a rise in household income will not increase market demand for all goods and services. The effect of a rise in income on demand for an individual good will depend on the nature of the good.

Demand and the level of increase may be related in different ways (see Figure 4).

- (a) A rise in household income may increase demand for a good. This is what we might normally expect to happen, and goods for which demand rises as household income gets bigger are called normal goods.
- (b) Demand may rise with income up to a certain point but then falls as income rises beyond that point. Goods whose demand eventually falls as income rises are called inferior goods, e.g. tripe, cheap wine. The reason for falling demand is that as incomes rise, demand switches to superior products e.g. beef instead of tripe, better quality wines instead of a cheaper variety.



The response of demand for a good to a change in the consumer's income is indicated by the income elasticity of demand. This will be discussed further in the next chapter.

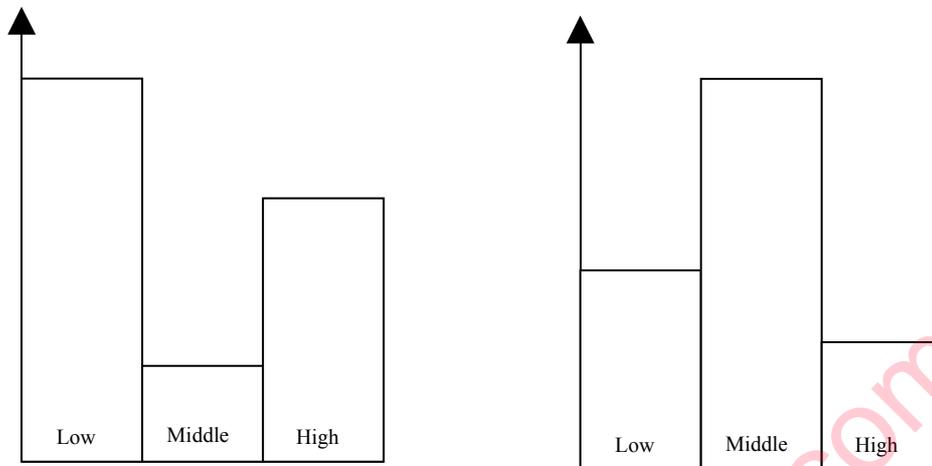
Tastes, fashion and expectations

A change in fashion will alter the demand for a product. For example, if it becomes fashionable for middle class households in the UK to drink wine with their meals, expenditure on wine will increase. There may be passing 'crazes', such as roller skates or skateboards.

If consumers believe that prices will rise, or that shortages will occur, they may attempt to stock up on the product, thereby creating excess demand in the short term which will increase prices. This can then lead to panic buying. Examples include fear of war, the budget, the effect of strikes or a rumour.

Market demand and the distribution of national income

Market demand for a good is influenced by the way in which the national income is shared between households. Consider the following patterns of income distribution:



In Figure 5(a), which has many rich and poor households and few middle income ones, there should be a large demand for luxury cars and yachts and also for staple foods such as bread and potatoes. In case (b), there should be high demand for medium-sized cars and TV sets, and other 'middle income' goods.

To summarise, the market demand curve relates the total quantity of a commodity demanded to its own price, on the assumption that all other prices, household incomes, the distribution of total income among households, assets and expectations and other factors are held constant.

Changes in demand

If the price of a good goes up or down, given no changes in the other factors that directly affect price, then there will be a shift in the Quantity demanded along the demand curve.

When there is a change in other factors that affect demand, the relationship between demand quantity and price will also change, and there will be a different price/quantity demand schedule and so a different demand curve. We refer to these changes as a shift of the demand curve.

This is an important distinction, which bears repetition in a slightly different form to consolidate your understanding of it.

- (a) Movements along a demand curve for a good are caused by changes in its price
- (b) Shifts in the demand curve for a good are caused by any of the other factors which affect demand for a good, other than its price.

Shifts of the demand curve

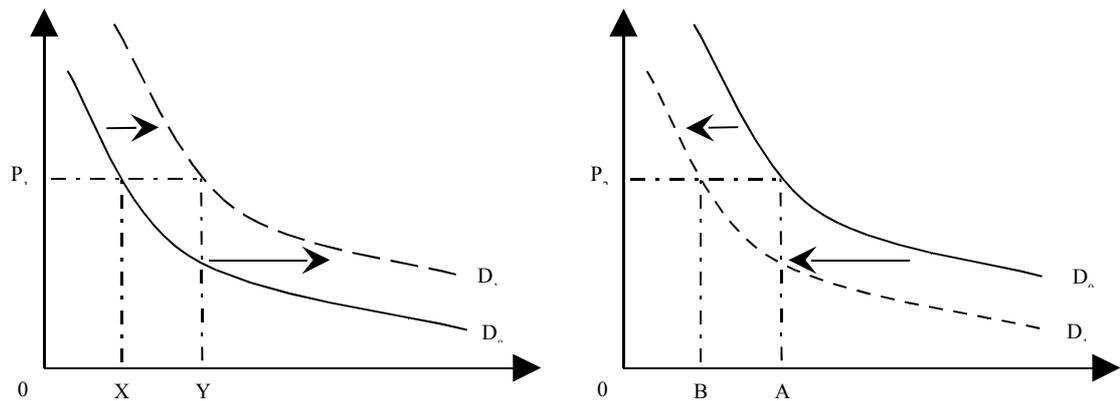


Figure 6(a) depicts a rise in demand at each price level, with the demand curve shifting to the right, from D_0 to D_1 . For example, at price P_1 , demand for the good would rise from X to Y . This shift could be caused by any of the following:

- a rise in household income;
- a rise in the price of substitutes;
- a fall in the price of complements;
- a change in tastes towards this product;
- an expected rise in the price of the product. If the price rise later fails to occur, the shift in demand would be temporary. If the price rise does occur, the demand curve would revert to its previous position, but fewer goods would now be demanded because the price is now higher.

Figure 6(b) depicts a fall in demand at each price level which is represented by a shifting to the left of the demand curve, from 'old' curve D_0 to 'new' curve D_1 . This shift may be caused by the opposite of the changes described in the previous paragraph

For example, at price P_2 the demand will fall from A to B .

To summarise:

- a shift of the demand curve to the right portrays an increase in the quantity demanded at any given price level;
- a shift of the curve to the left portrays a reduction in the Quantity demanded at any given price level.

Supply

The concept of supply

Supply refers to the quantity of a good that existing suppliers or would-be suppliers would want to produce for the market at a given price. In this context, and usually in economics, the term 'good' includes services as well as tangible products.

The quantity of a good that can be supplied to a market varies up or down, either because:

- (a) existing suppliers increase or reduce their output quantities, or
- (b) firms stop producing altogether and leave the market, or new firms enter the market and begin producing the good.

If the Quantity that firms want to produce at a given price exceeds the Quantity that purchasers would demand, there would be an excess of supply, with firms competing to win what sales demand there is. Over-supply and competition would then be expected to result in price-competitiveness and a fall in prices.

As with demand, supply relates to a period of time -e.g. an annual rate of supply quantities or a monthly rate.

As with demand, a distinction should be made between:

- (a) market supply, which is the total Quantity of the good that all firms in the market would want to supply at a given price; and
- (b) an individual firm's supply schedule which is the quantity of the good that the individual firm would want to supply to the market at any given price.

Factors which influence the supply quantity

The quantity supplied of a good depends on:

- (a) the price obtainable for the good;
- (b) the prices of other goods. An increase in the price of other goods would make the supply of a good whose price does not rise less attractive to suppliers;
- (c) the cost of making the good, which in turn depends on the prices of factors of production -ie wages, interest rates, land rents and profit expectations. A rise in the price of one factor of production (say labour), will cause producers to shift away from supplying goods whose costs and profits are closely related to the price of labour, towards the supply of goods where the cost of labour is less significant;
- (d) changes in technology. Technological developments which reduce costs of production (and increase productivity) will raise the quantity of supply of a good;
- (e) changes in the weather may affect the availability of supply (e.g. agricultural goods).

The factors that affect supply can be summarised briefly as prices and costs, and so profits. In this chapter, our main interest is with the influence of price on supply and demand. The influence of cost on output decisions by firms, and so on supply to the market, will be discussed more fully in a later chapter.

Supply and the price of a good

In general, suppliers will want to supply a greater quantity of their output at higher prices. For example, if the price of product X is £5 per unit, a supplier might be willing to supply 1,000 units of the product to consumers in the market at that price. If the price of product X now goes up to £10, the supplier will be willing to supply more than 1,000 units of the product. Just how many more than 1,000 he would want to supply would depend on circumstances.

Why would a supplier be willing to supply more output at a higher price?

It might seem logical to suppose that higher prices should mean higher profits, and so the firm would be attracted by the prospect of bigger profits into supplying more units of output. This is not the full answer, though. We must also ask why, in our example, the supplier was only willing to supply 1,000 units of product X at a price of £5.

The answer must presumably be that it would not be worthwhile, and that the unit cost of making extra output would exceed the sales price of £5 per unit. This might be, for example, because the supplier's capacity is already fully extended, and the costs of increasing capacity (anything from overtime premiums to the cost of a new factory) would not be covered by the £5 per unit sales price. At a higher selling price per unit, the output limit where unit costs begin to exceed unit prices will be at a higher level, and so the supplier would now be willing to produce more at the new higher price.

The supply curve

A supply schedule and supply curve can be drawn:

- (a) for an individual supplier; or
- (b) for all firms which produce the good. market supply curve. This total supply curve of all suppliers is the

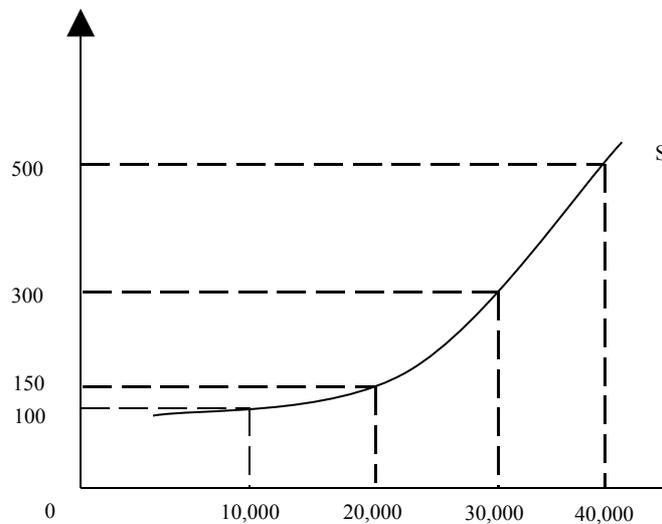
Example

Suppose that the supply schedule for product y is as follows.

<i>Price per unit</i>	<i>Quantity that suppliers would supply at this price</i>
£	Units
100	10,000
150	20,000
300	30,000
500	40,000

A supply curve is constructed in a similar manner to a demand curve (from a schedule of supply quantities at different prices) but shows the quantity suppliers are willing to produce at different price levels. It is an upward-sloping curve from left to right, because greater quantities will be produced at higher prices.

Contrast the inverse relationship between price and Quantity demanded which we examined in our analysis of demand. In the case of supply, the relationship between price and Quantity supplied is positive: as price rises, so too does the Quantity that producers are willing to supply. The relationship between output and price, using the data in our example, is shown as a supply curve in Figure 7.



A supply curve shows how the quantity supplied will change in response to a change in price, provided that all other conditions affecting supply remain unchanged (*ceteris paribus*). If supply conditions (the price of other goods, or costs of factors of production, or changes in technology) alter, a different supply curve must be drawn. In other words, a change in price will cause a shift in supply along the supply curve. A change in other supply conditions will cause a shift in the supply curve itself.

This corresponds to the important point made earlier about demand curves, and again bears repetition.

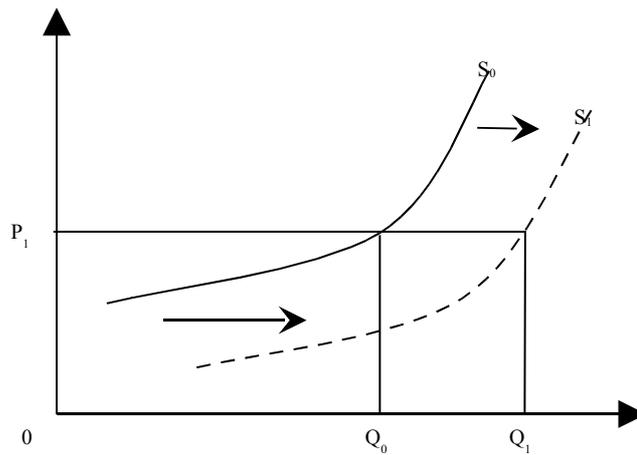
- (a) Movements along a supply curve represent changes in the total quantity of a good that suppliers would want to supply when there is a change in the price of the good.
- (b) Shifts of the supply curve represent changes in the total quantity of a good that suppliers would want to supply at all prices, because of a change in the cost of supply -eg technological progress, or a change in the price of other goods.

Shifts of the market supply curve

The market supply curve is the aggregate of the supply curves of individual firms in the market. A shift of the market supply curve occurs when supply conditions -i.e. factors influencing supply, other than the price of the good itself -alter (e.g. the price of factors of production, the prices of other goods, technology etc). Figure 8 shows a shift in the supply curve from S_0 to S_1 . A rightwards shift of the curve shows an expansion of supply and may be caused by:

- (a) a fall in the cost of factors of production;
- (b) a fall in the price of other goods. The production of other goods becomes relatively less attractive as their price falls. We therefore expect that (*ceteris paribus*) the supply of one good will rise as the prices of other goods fall (and vice versa);
- (c) technological progress -which reduces unit costs and also increases production capabilities.

In effect, a shift of the supply curve is the result of changes in costs, either in absolute terms or relative to the costs of other goods.



If the price of the good is P_1 (Figure 8), suppliers would be willing to increase supply from Q_0 to Q_1 under the new supply conditions.

Note that we need to distinguish between short run and long run responses of both supply and demand. In the short run both supply and demand are relatively unresponsive to changes in price, as compared to the long run.

In the case of supply, changes in the quantity of a good supplied often require the laying off or hiring of new workers, or the installation of new machinery. All of these changes, brought about by management decisions, must take some time to implement.

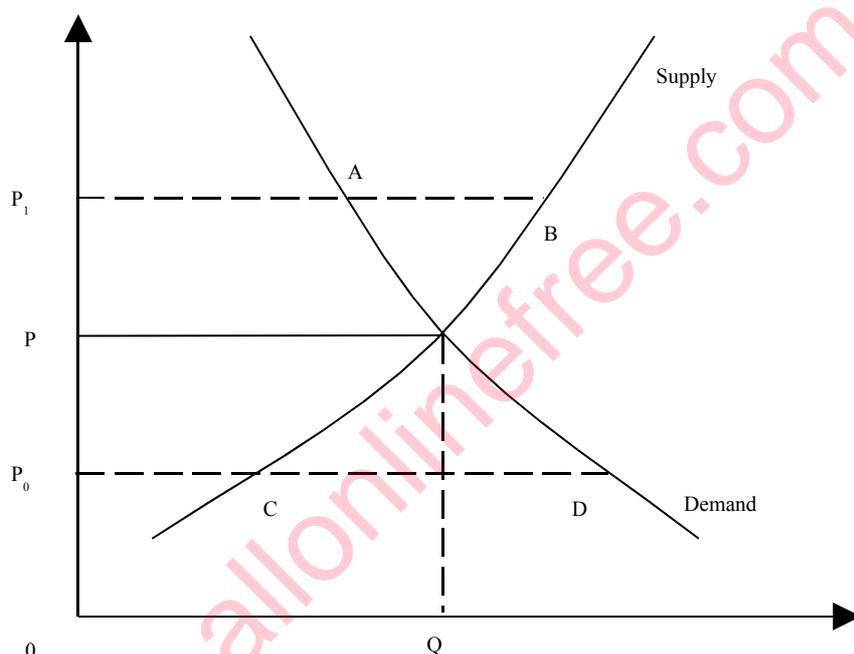
In the case of demand, it takes time for consumers to adjust their buying patterns, although demand will often respond more rapidly than supply to changes in price or other demand conditions. In some markets, such as the market for chocolate bars, responses to changes in price are relatively rapid. In others, such as the market for military aircraft, response times are much longer.

The price mechanism

If demand for a good exceeds supply, consumers must either stop demanding what they cannot have, or they must be prepared to pay more for the good. At a higher price, firms will be prepared to supply more of the good. On the other hand, if the price of a good is such that firms want to supply more than consumers are willing to buy, production must be cut back in volume or the price must be reduced so as to stimulate demand.

The equilibrium price

The price mechanism brings demand and supply into equilibrium and the equilibrium price for a good is the price at which the volume demanded by consumers and the volume that firms would be willing to supply are the same.



At price P_1 there is an excess of the quantity that suppliers want to produce over the quantity demanded at that price, equal to the distance AB. The reaction of suppliers as unsold stocks accumulate would be to:

- cut down the current level of production in order to disaccumulate unwanted stocks (i.e. de-stock); and/or
- reduce prices in order to encourage sales.

The opposite will happen at price p_0 where there is an excess of demand over supply shown by the distance CD. Output and price would increase.

At price p the amount that sellers are willing to supply is equal to the amount that customers are willing to buy. There will be no unusual variation in stocks and, unless something else changes, there will be no change in price. p is the equilibrium price.

At the equilibrium price P , consumers will be willing to spend a total of $P \times Q$ -i.e. PQ - on buying Q units of the product, and suppliers will be willing to supply Q units to earn revenue of $P \times Q$. The forces of supply and demand push a market to its equilibrium price and quantity.

- (a) If there is no change in conditions of supply or demand, the equilibrium price will rule the market and will remain stable.
- (b) If the equilibrium price does not rule, the market is in disequilibrium, but supply and demand will push prices towards the equilibrium price.
- (c) Shifts in the supply curve or demand curve will change the equilibrium price and quantity.

The 'law' of supply and demand is that in a free market, the equilibrium price and output level of a good is the price and output level at which the market demand curve and the market supply curve intersect.

In a free market, price acts as a mechanism which signals demand and supply conditions to producers and consumers. It therefore determines the activities of both producers and consumers, influencing the levels of demand for and the supply of goods.

'The price system was not consciously created. It does not require that anyone consciously foresee and co-ordinate the necessary changes; adjustments occur automatically as a result of the separate decisions taken by a large number of individuals, all seeking their own best interest.' (*R G Lipsey*)

www.allonlinefree.com

Elasticity Of Demand And Supply

Your objectives

After completing this chapter you should:

- (a) understand the concept of elasticity in relation to both supply and demand;
- (b) be aware of some of the factors that influence elasticity of supply and demand in the cases of particular products.

Price elasticity of demand

The concept of elasticity

So far we have discussed the direction of changes in demand and supply when prices change. When price goes up, the quantity demanded will fall, and the quantity suppliers will be willing to produce will go up. But we have not yet considered the extent of these changes. For example, if prices went up by 10%, would the quantity demanded fall by 5%, 20%, 50% or what? And would the quantities that suppliers would want to produce go up by 5%, 10%, 15% or what? The extent of these changes is measured by the price elasticity of demand and the elasticity of supply.

We have also discussed shifts in the demand curve due to changes in the price of substitutes and complements and changes in household income, but again we have not yet considered the extent of the shifts in demand arising from shifts in the demand curve. The extent of these changes is measured by:

- (a) the cross elasticity of demand, in the case of substitutes and complements;
- (b) the income elasticity of demand, in the case of household income.

The price elasticity of demand

Price elasticity of demand (often denoted by the Greek symbol η , pronounced 'eeta') is a measure of the extent of change in market demand for a good in response to a change in its price. It is measured as:

$$\eta = \frac{\text{change in Quantity demanded, as a percentage of demand}}{\text{change in price, as a percentage of the price}}$$

Since the demand goes up when the price falls, and goes down when the price rises, the elasticity has a negative value, but it is usual to ignore the minus sign.

Elastic and inelastic demand

As mentioned already, the elasticity of demand will generally have a negative value since demand curves are normally downward sloping. However, since the minus sign is often ignored, an elasticity of -1 is usually referred to as elasticity of 1, or unity. The value of demand elasticity may be anything from zero to infinity.

Demand is:

- (a) inelastic if the absolute value is less than 1; and
- (b) elastic if the absolute value is greater than 1.

Where demand is inelastic, the Quantity demanded falls by a smaller percentage than price, and where demand is elastic, demand falls by a larger percentage than the percentage rise in price.

Price elasticity and the slope of the demand curve

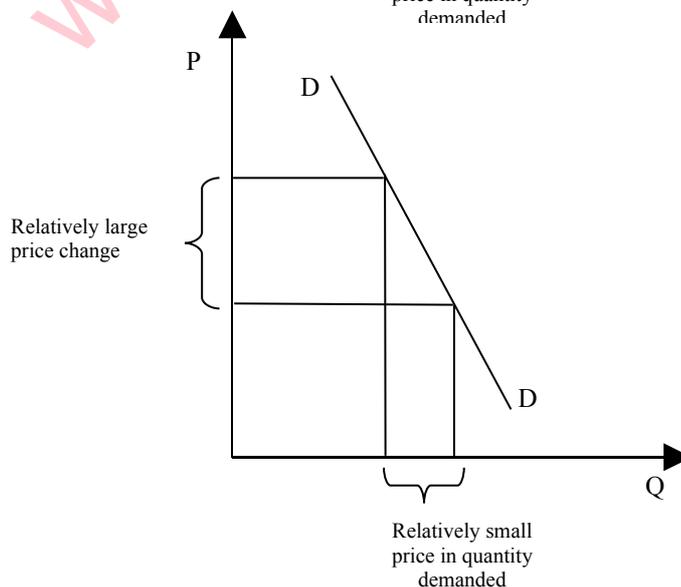
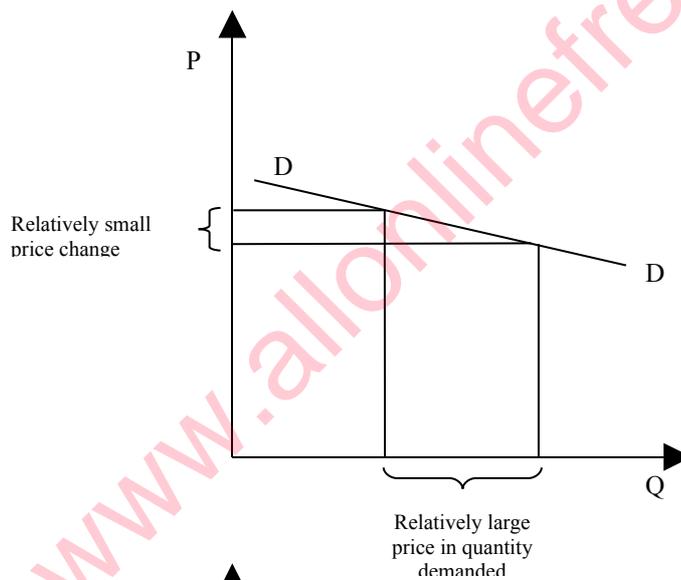
Generally, demand curves slope downwards. Consumers are willing to buy more at lower prices than at higher prices. Except in certain cases (which are referred to later), elasticity will vary in value along the length of a demand curve.

At 'high' prices (the top of the demand curve), small percentage price reductions can bring large percentage increases in quantity demanded. This means that demand is elastic over these ranges, and price reductions bring increases in total expenditure by consumers on the commodity in question.

At 'lower' prices (the bottom of the demand curve), large percentage price reductions can bring small percentage increases in quantity. This means that demand is inelastic over these price ranges, and price increases result in increases in total expenditure.

Elasticity is rarely the same at all price levels. For example, the price elasticity of demand for a good might be elastic (greater than 1) at some prices and inelastic (less than 1) at other prices.

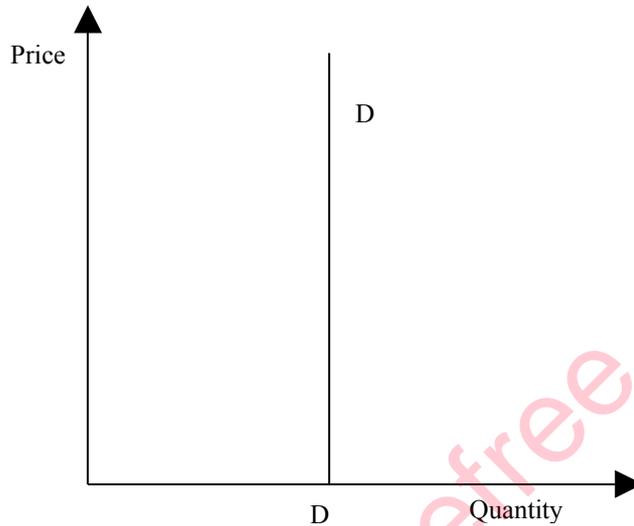
It is not possible merely by looking at the slopes of any two curves to state their comparative elasticity's over other, different price ranges, even if there is the same absolute fall in price. This is because the elasticity of a demand curve is not constant over the entire curve. However, it is possible to say that if a demand curve becomes steeper over a particular range of quantity, then demand is becoming more inelastic. Conversely, a shallower demand curve over a particular range indicates more elastic demand. This is illustrated in Figure 1 (a) and 1 (b) below.



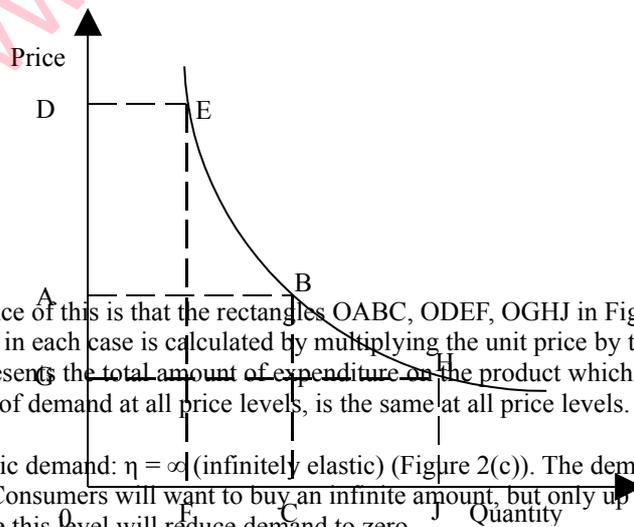
Special values of price elasticity

There are three special values of price elasticity of demand; 0, 1 and infinity.

$\eta = 0$. Demand is perfectly inelastic (Figure 2(a)). The demand curve is a vertical straight line and there is no change in quantity demanded, regardless of the change in price. A close example to this in real life is the demand for table salt which will be largely unaffected by price changes; we would not expect people to demand more salt if the price halved nor less salt if the price doubled.

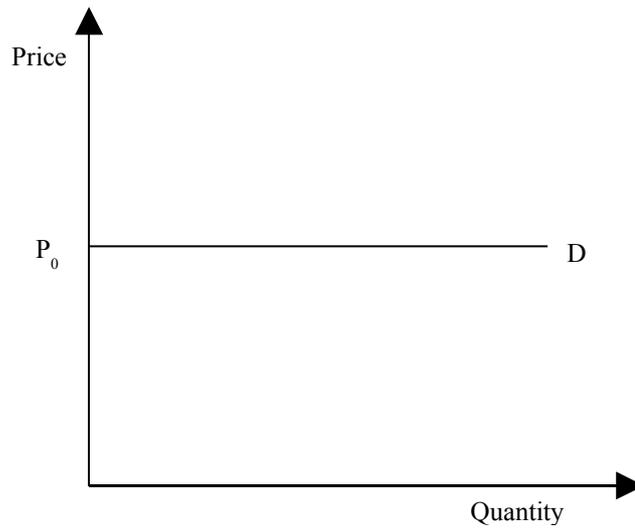


Unit elasticity of demand: $\eta = 1$ (Figure 2(b)). The demand curve of a good whose elasticity is lower over its entire range is a rectangular hyperbola.

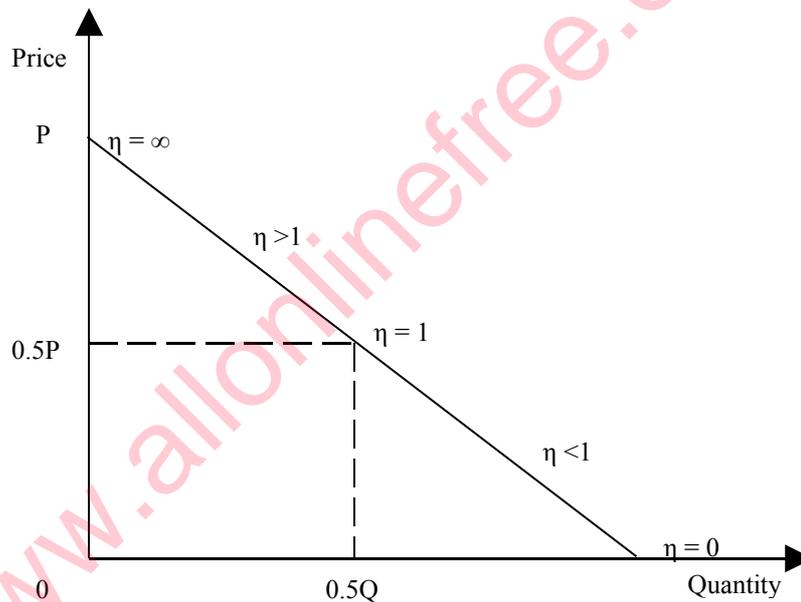


The significance of this is that the rectangles OABC, ODEF, OGHJ in Figure 2(b) all have the same area. The area in each case is calculated by multiplying the unit price by the quantity sold: in other words, it represents the total amount of expenditure on the product which, in the case of a product with unit elasticity of demand at all price levels, is the same at all price levels.

Perfectly elastic demand: $\eta = \infty$ (infinitely elastic) (Figure 2(c)). The demand curve is a horizontal straight line. Consumers will want to buy an infinite amount, but only up to price level P_0 . Any price increase above this level will reduce demand to zero.



The ranges of price elasticity at different points on a downward-sloping straight line demand curve are illustrated in Figure 3.



By studying Figure 3 you may be able to see the justification for the following important rules.

- (a) For a commodity whose price elasticity is greater than one over the range of the price change:
 - (i) a reduction in price produces an increase in total expenditure on the commodity;
 - (ii) a rise in price produces a reduction in total expenditure on the commodity.
- (b) For a good whose price elasticity is less than
 - (i) a reduction in price produces a reduction in total expenditure on the commodity;
 - (ii) an increase in price produces an increase in total expenditure on the commodity.

Business people can make use of information on how consumers will react to pricing decisions, not least because of the effect of this on profits. Government policy makers can use information about elasticity when making decisions about indirect taxation. Items with a low price elasticity of demand, such as cigarettes and alcohol, tend to be targets for high taxation since by increasing taxes on these, total revenue can be increased. If cigarettes were price elastic, increases in taxation would be counter-productive as they would result in lower government revenue.

Factors influencing price elasticity of demand for a good

Factors that determine price elasticity of demand are similar to the factors other than price that affect the volume of demand. Elasticity is really a measure of the strength of these other factors on demand. The main factors affecting price elasticity of demand are:

- (a) the availability of substitutes;
- (b) the time horizon;
- (c) the pricing policies of competitors.

The availability of close substitutes. The more substitute goods there are, especially close substitutes, the more elastic will be the price elasticity of demand for a good. For example, in a greengrocer's shop, a rise in the price of one vegetable such as carrots or cucumbers is likely to result in a switch of customer demand to other vegetables, many vegetables being fairly close substitutes for each other. Again the elasticity of demand for a particular brand of breakfast cereals is much greater than the elasticity of demand for breakfast cereals as a whole, because the former have much closer substitutes. This factor is probably the most important influence on price elasticity of demand.

The time period. Over time, consumers' demand patterns are likely to change. If the price of a good is increased, the initial response might be very little change in demand (inelastic demand) but then as consumers adjust their buying habits in response to the price increase, demand might fall substantially. The time horizon influences elasticity largely because the longer the period of time which we consider, the greater the knowledge of substitution possibilities by consumers and the provision of substitutes by producers.

Competitors' policies. If the response of competitors to a price increase by one firm is to keep their prices unchanged, the firm raising its prices is likely to face elastic demand for its goods at higher prices. If the response of competitors to a reduction in price by one firm is to match the price reduction themselves, the firm is likely to face inelastic demand at lower prices. This is a situation which probably faces many large firms with one or two major competitors (i.e. oligopolies).

Other factors influencing elasticity of demand. Generally speaking, the higher proportion of a consumer's income that a commodity takes up, the greater will be the elasticity of demand. Thus motor vehicle demand can be expected to be more responsive to a 10% price change than, say, the demand for postage stamps or for milk.

Other elasticities of demand

Income elasticity of demand

The income elasticity of demand for a good indicates the responsiveness of demand to changes in household incomes. It is measured as:

$$\frac{\text{change in quantity demanded, as a percentage of demand}}{\text{change in household income, as a percentage of household income}}$$

(a) A good is income elastic if income elasticity is greater than 1 so that quantity demanded rises by a larger percentage than the rise in income. For example, if the demand for compact discs will rise by 10% if household income rises by 7%, we would say that compact discs are income elastic.

(b) A good is income inelastic if income elasticity is between 0 and 1 and the quantity demanded rises less than the proportionate increase in income. For example, if the demand for books will rise by 6% if household income rises by 10%, we would say that books are income inelastic.

Both of these categories are said to be normal goods, which means that demand for them will rise when household income rises, and so they have a positive income elasticity of demand.

If income elasticity is less than 0, income elasticity is negative and the commodity is said to be an inferior good since demand for it falls as income rises.

Cross elasticity of demand

Cross elasticity of demand refers to the responsiveness of demand for one good to changes in the price of another good.

$$\text{Cross elasticity of demand} = \frac{\text{change in quantity of good A demanded, as a percentage of quantity demanded*}}{\text{change in the price of good B, as a percentage of price}}$$

*(given no change in the price of A)

- (a) If the two goods are substitutes, cross elasticity will be greater than 0 and a fall in the price of one will reduce the amount demanded of the other.
- (b) if the goods are complements, cross elasticity will be negative and a fall in the price of one will raise demand for the other .

Cross elasticity involves a comparison between two products. Cross elasticity is significant where the two goods are close substitutes for each other, so that a rise in the price of B, say, is likely to result in an increase in the demand for A.

Cross elasticity of demand between two complementary products can also be significant because a rise in the price of B would result in some fall in demand for A because of the fall in demand for B.

Demand elasticity and time

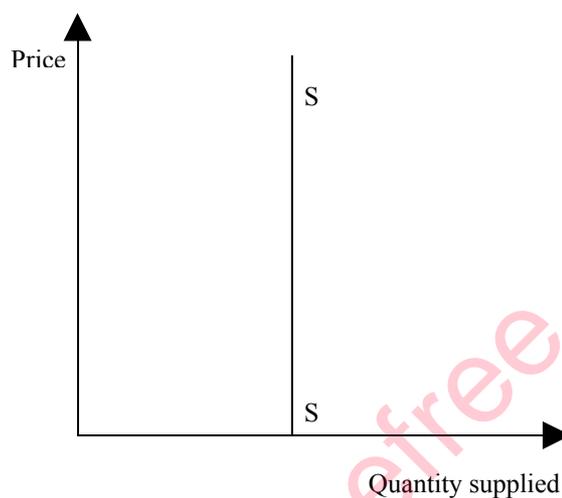
Elasticity of demand for any good -price elasticity, income elasticity and cross elasticity -can and does change over time. Generally, there are some substitutes for all goods.

Elasticity of supply

The elasticity of supply indicates the responsiveness of supply to a change in price. It is measured as:

$$\frac{\text{change in Quantity supplied, as a percentage of quantity supplied}}{\text{change in price, as a percentage of price}}$$

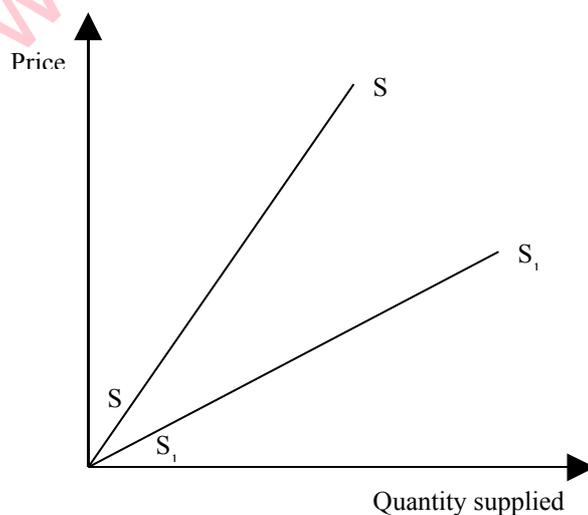
The elasticity of supply will vary between the three illustrated in Figure 4 (a), (b) and (c)



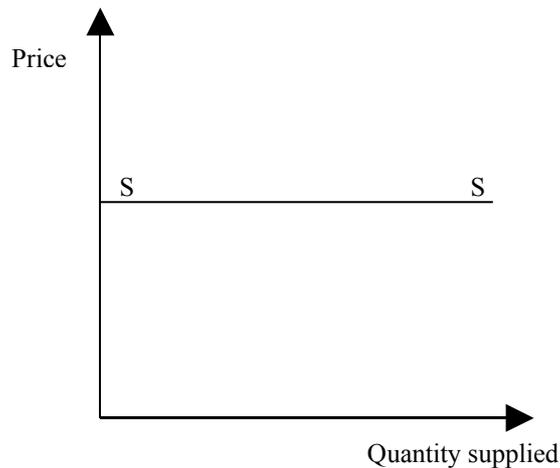
Where the supply of goods is fixed whatever price is offered, e.g. in the case of antiques, vintage wines and land, supply is perfectly inelastic and the elasticity of supply is 0 (Figure 4(a)).

Where the supply of goods varies proportionately with the price, elasticity of supply equals one (Figure 4(b)). Both supply curves in the following diagram have unit elasticity because they are straight line curves passing through the origin.

(Note that a demand curve with unit elasticity along all of its length is not a line, but a supply curve with unit elasticity is a straight line.)



Where the producers will supply any amount at a given price but none at all at a slightly lower price, elasticity of supply is infinite, or perfectly elastic (Figure 4(c)).



Supply is elastic when the percentage change in the amount producers want to supply exceeds the percentage increase/decrease in price. Supply is inelastic when the amount producers want to supply changes by a smaller percentage than the percentage change in price.

Elasticity of supply and time

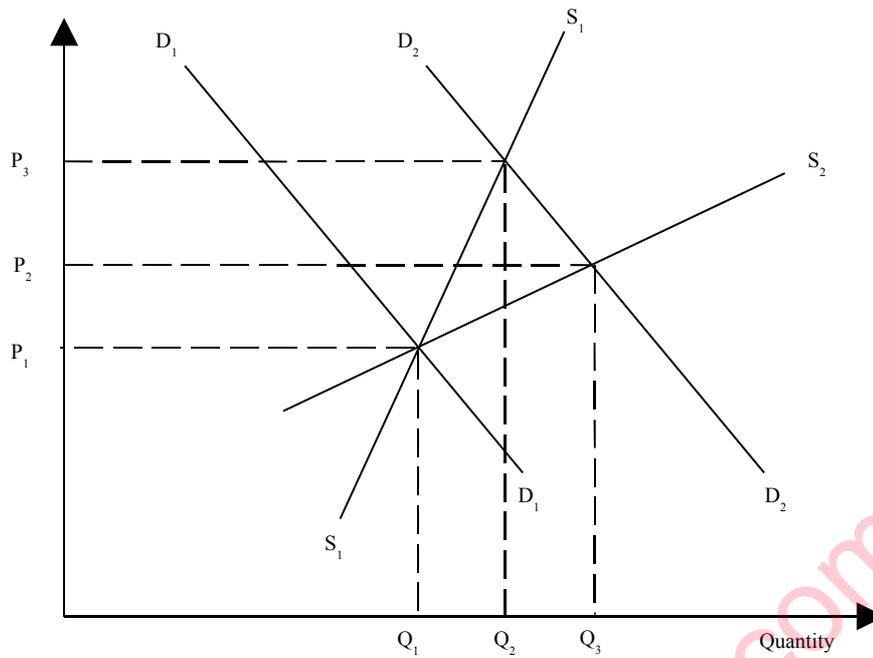
As with elasticity of demand, the elasticity of supply of a product varies according to the time period over which it is measured. For analytical purposes, four lengths of time period may be considered:

- (a) the market period which is so short that supplies of the commodity in question are limited to existing stocks. In effect, supply is fixed;
- (b) the short run is a period long enough for supplies of the commodity to be altered by increases or decreases in current output, but not long enough for the fixed equipment (ie plant, machinery, etc) used in production to be altered. This means that suppliers can produce larger quantities only if they are not already operating at full capacity; they can reduce output fairly quickly by means of lay-offs and redundancies. A common way of referring to this is to say that the short run is a period during which at least one of the factors of production is fixed in supply;
- (c) the long run is a period sufficiently long to allow firms' fixed equipment to be altered. There is time to build new factories and machines, and time for old ones to be closed down. New firms can enter the industry in the long run;
- (d) the secular period is so long that underlying economic factors such as population growth, supplies of raw materials (such as oil) and the general conditions of capital supply may alter. ('Secular' is derived from the Latin word 'saecula' meaning 'centuries'.) The secular period is ignored by economists except in the theory of economic growth.

These types of time period were postulated by Alfred Marshall in his Principles of economics (1920).

The price elasticity of supply can be seen as a measure of the readiness with which an industry responds to a change in price following a shift in the demand curve. An example will help to show what this means.

Suppose that there is an increase in the demand for restaurant meals in a city, shown by the rightward shift in the demand curve in Figure 5 below from D_1 to D_2 . The capacity of the industry is limited in the short run by the number of restaurants in operation. The restaurants can be used more intensively to a certain extent, and so supply is not perfectly inelastic, but supply is relatively inelastic because of the limit to this process. As a result, in the short run there is a large increase in the price from P_1 to P_2 .



The rise in price in the short run will encourage entrepreneurs to open new restaurants to take advantage of the profits to be earned. In the long run, supply is consequently more elastic and is shown by supply curve S_2 . The expanded output in the industry leads to a new equilibrium at a lower price P_3 with the new level of output being Q_3 .

Elasticity of supply and the cost of factors of production

The cost of factors of production will affect the elasticity of supply. If demand increases, the supplier might attract more factors of production. If the cost of attracting new amounts of the factor is high, the supply curve will be more inelastic than if this cost is low, because suppliers will need a higher price rise to cover these costs, in order to justify an increase in supply.

Market Failure: Externalities And Intervention

Your objectives

After completing this chapter:

- (a) be aware of the reasons why free markets may not lead to an ideal allocation
- (b) understand the concepts externalities, public goods and merit goods
- (c) understand why, how and for what purpose a government may intervene in a market in order to improve the allocation of resources

Market imperfections and market failure

Introduction

In this chapter, we shall be concerned with why a free market would result in an allocation of resources that is not optimal. If a free market has certain weaknesses and drawbacks, the question arises of whether there ought to be some regulation of the markets by the authorities (ie a mixed economy), in order to improve the allocation of resources.

In the case for a free market

Advocates of a free market economy argue that the market forces of supply and demand will result in an efficient allocation of economic resources.

- (a) Consumers will decide what they want to buy by relating the prices of goods to their marginal utilities.
- (b) Producers will decide what goods to produce, and in what quantities, by relating their prices to the costs of production (and the costs of the scarce resources needed to produce them).

Market failure

Market failure refers to a situation in which the market mechanism fails to produce the best (the 'optimal') allocation of resources.

Market failure is caused by a number of factors, which might be listed as:

- (a) imperfections in the market;
- (b) a divergence between private costs and social costs (externalities);
- (c) public goods;
- (d) the need to consider non-market goals, such as 'social justice'.

Externalities and public goods

Social costs and private costs

In a free market, suppliers and households make their output and buying decisions for their own private benefit, and these decisions determine how the economy's scarce resources will be allocated to production and consumption. Private costs and private benefits therefore determine what goods are made and bought in a free market.

It can be argued that a free market mechanism would result in a satisfactory allocation of resources, provided that private costs are the same as social costs, and private benefits are the same as social benefits.

When private benefit is not the same as social benefit, or when private cost is not the same as social cost, an allocation of resources, which reflects private costs and benefits only may not be socially acceptable.

Private cost measures the cost to a firm of the resources it uses to produce a good. (This economic cost, as we shall see in a later chapter, is the opportunity cost of the resources used).

Social cost measures the cost to society as a whole of the resources that a firm uses

An example where private cost and social cost differ is where a firm produces a good, and during the production process, pollution is discharged into the air. The private cost to the firm is the cost of the resources needed to make the good. The social cost consists of the private cost plus the additional 'costs' incurred by other members of society, who suffer from the pollution.

Private benefit measures the benefit obtained directly by a supplier or by a consumer .

Social benefit measures the total benefit obtained, both directly by a supplier or a consumer, and indirectly, at no extra cost, to other suppliers or consumers.

An example where private benefit and social benefit differ is where customers at a cafe in a piazza in Italy benefit from the entertainment provided by professional musicians, who are hired by the cafe. The customers of the cafe are paying for the service, in the prices they pay, and they obtain a private benefit from it. At the same time, other people in the piazza, who are not customers of the cafe, might stop and listen to the music. They will obtain a benefit, but at no extra cost to themselves. The social benefit from the musicians' service is greater than the private benefit to the cafe's customers.

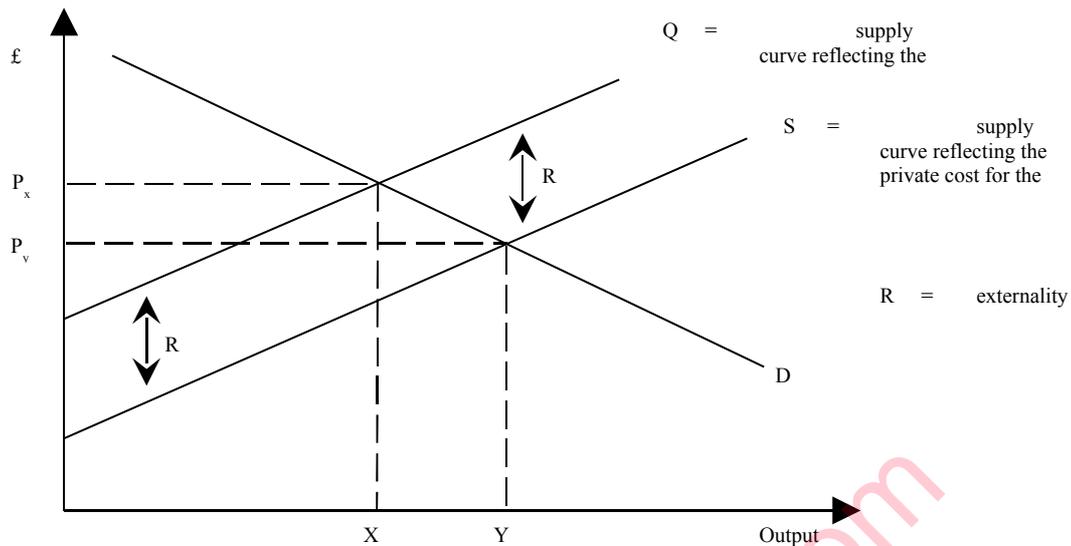
Externalities

An 'externality' may be defined as a difference between the private and the social costs, or benefits, arising from an activity. Less formally, an 'externality' is a cost or benefit which the market mechanism fails to take into account because the market responds to purely private signals.

The consequence of externalities is market failure to achieve a socially satisfactory allocation of resources, which properly reflects social costs and benefits.

We can use demand and supply analysis to illustrate the consequences of externalities. However, you need to accept that if an adverse externality exists, so that the social cost of supplying a good is greater than the private cost to the supplier firm, then:

- (a) a supply curve which reflects total social costs will be to the left of the (private cost) market supply curve; and
- (b) the vertical distance between the private cost supply curve and the social cost supply curve represents the size of the externalities



- (a) if a free market exists, the amount of the good produced will be determined by the interaction of demand (curve D) and supply curve S. Here, output would be Y, at price P_y ;
- (b) if social costs were taken into account, and the market operated successfully, the amount of the good produced should be just X, at price P_x .

Given a free market, output of the good will exceed what it ideally should be, and so resources will have been over-allocated to production of this particular good.

Public goods

Some goods, by their very nature, involve so much 'spillover' of externalities that they are difficult to provide except as public goods whose production is organised by the government.

In the case of such goods, the consumption of the good by one individual or group does not significantly reduce the amount available for others. And if one individual makes use of the good, it does not reduce the availability of the good and its benefits to other individuals. Furthermore, it is often difficult or impossible to exclude anyone from its benefits, once the good has been provided.

Examples of such goods are:

- (a) defence and the armed forces;
- (b) the police service;
- (c) the fire service;
- (d) the Thames flood barrier in England, and the Aswan Dam in Egypt.

Defence and policing are perhaps the most obvious examples of 'naturally' public goods. It is not practicable for individuals to buy their own defence systems or policing arrangements.

Merit goods and demerit goods

There are some types of goods which, in the absence of intervention, the market would not supply at all, or would supply in inappropriate quantities. This is because individuals do not always realise where their best interests lie, and so might fail to purchase goods which would benefit them or, alternatively, might purchase goods which do not benefit them.

A government may intervene to rid the market of this perceived inefficiency.

- (a) The government encourages, or insists upon, the purchase of goods (merit goods) which it believes consumers should buy. For example, if you use a motorbike you are compelled by law to purchase a helmet.
- (b) The government discourages, or prohibits, the purchase of goods (demerit goods) which it believes should not be bought by consumers. For example, children are not allowed to purchase cigarettes, and nobody (other than medical practitioners etc) is allowed to purchase addictive drugs.

Price regulation

Market failure and the case for regulation of the market

Market failure, remember, refers to the failure of a free market to result in an optimal allocation of resources.

The existence of market failure and of externalities suggests the need for regulation of the market by the government, in order to improve the allocation of resources.

The main ways in which a government might choose to regulate or control markets are as follows:

- (a) control the means of production (i.e. have state-owned industries);
- (b) provide public goods;
- (c) provide some goods (e.g. education, health) in greater quantities than there would be if an entirely free market operated;
- (d) influence markets through legislation and regulation (e.g. regulation of monopolies, bans on dangerous drugs, enforcement of the use of some goods such as car seat belts, laws on pollution control, and so on);
- (e) re-distribute wealth, perhaps by taxing relatively wealthy members of society and re-distributing this tax income so as to benefit the poorer members;
- (f) influence market supply and demand through:
 - (i) price legislation;
 - (ii) indirect taxation -for example, the lower tax on lead-free petrol in the UK compared with leaded petrol is aimed at encouraging greater demand for unleaded petrol so as to reduce environmental pollution;
 - (iii) subsidies;
- (g) intervene in the market to create a demand for output that is labour-creating. A free price mechanism will result in a total demand for goods and services that would be met by a matching total supply, but this total supply quantity might be insufficient to create full employment within the economy. Government might therefore wish to intervene to create a demand for output in order to create more jobs.

Price regulation

The government might introduce regulations either:

- (a) to set a maximum price for a good, perhaps as part of an anti-inflationary economic policy (e.g. a prices and incomes policy); or

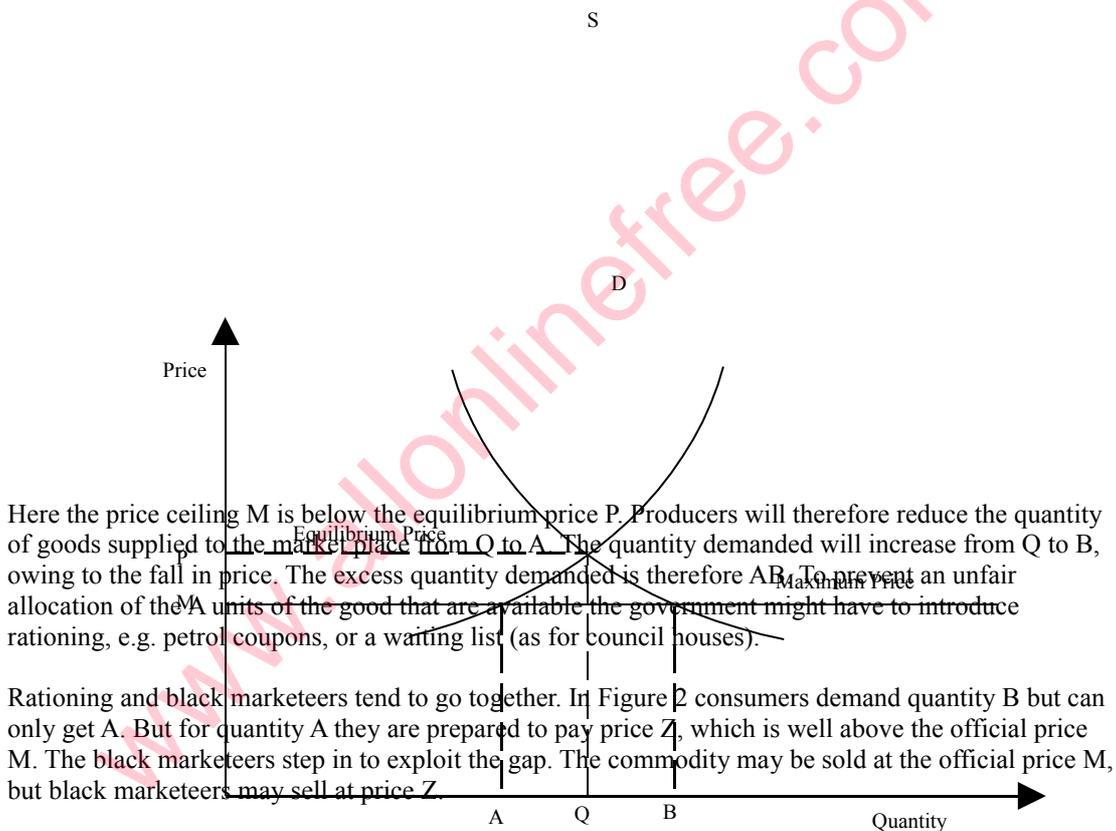
- (b) to set a minimum price for a good. For example, OPEC (the Organisation of Petroleum Exporting Countries) has tried in the past to impose minimum prices for oil on the world markets.

Maximum price legislation

The government may try to prevent prices of goods rising by establishing a price ceiling. If the price ceiling is higher than the equilibrium price, setting a price ceiling will have no effect at all on the operation of market forces.

The price ceiling is above the equilibrium market price and is therefore ineffectual because the equilibrium price and output is reached by the free market interaction of supply and demand.

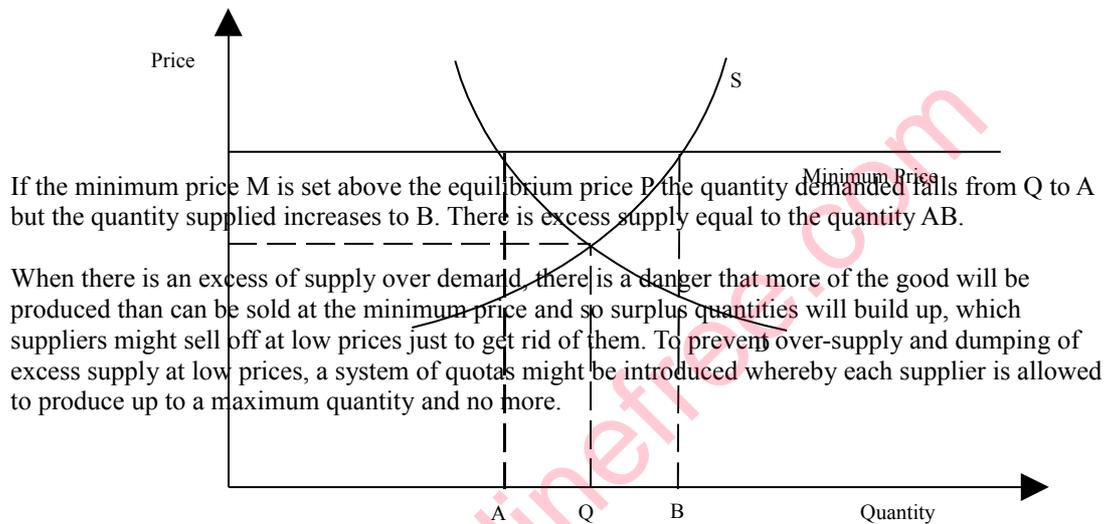
If the maximum price M is lower than what the equilibrium price would be, there will be an excess of demand over supply (Figure 2). The low price attracts customers, but deters suppliers.



Minimum price regulations

Minimum price legislation or regulations aim to ensure that suppliers earn a minimum price for each unit of output they sell. Minimum prices are used in the European Community (EC), and farmers are guaranteed a minimum price for their output by the authorities if they cannot get a higher price on the free market.

If the minimum price is set below the market equilibrium there is no effect. But if it is set above the market price, it will cause an excess supply of AB (as in Figure 3). This has been a recurring problem for the EC, resulting in the so-called 'butter mountains' and 'wine lakes'.



www.allonlinefranchise.com

Indirect taxes and subsidies

Indirect taxation

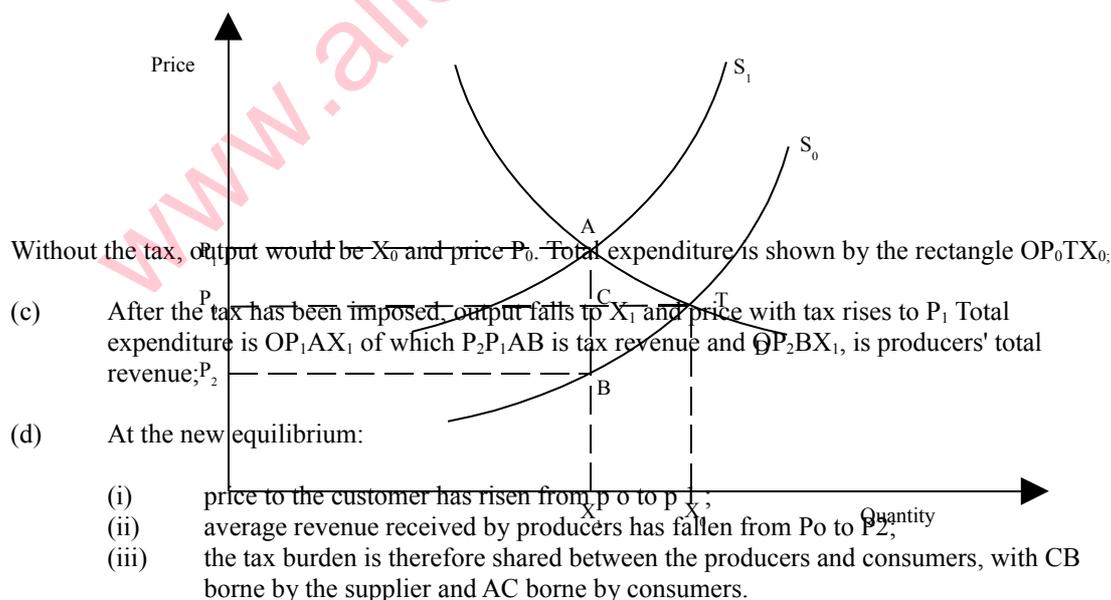
An alternative form of price and output regulation is indirect taxation. (i.e. taxation on goods or services as opposed to direct taxation on incomes)

An indirect tax which is imposed on some goods but not on others is called a selective indirect tax.

If an indirect tax is imposed on one good, the tax will shift the supply curve to the left. This is because the price to consumers includes the tax, but the suppliers still only receive the net-of-tax price. For example, in Figure 4:

- (a) the supply curve net of tax is S_0 ;
- (b) the supply curve including the cost of the tax is S_1 .

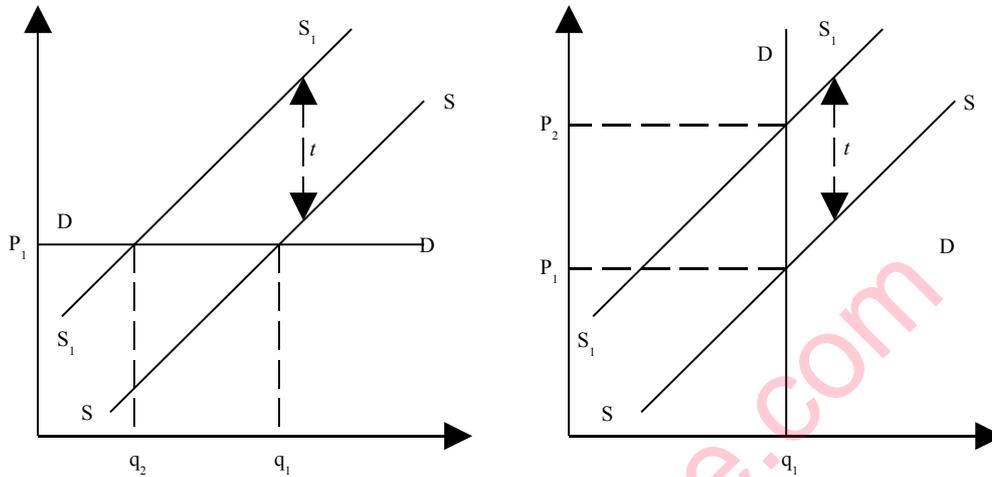
So if demand is for X_1 units, say, the price to suppliers will be P_2 but the price with tax to the consumer would be P_1 and the tax would be $(P_1 - P_2)$.



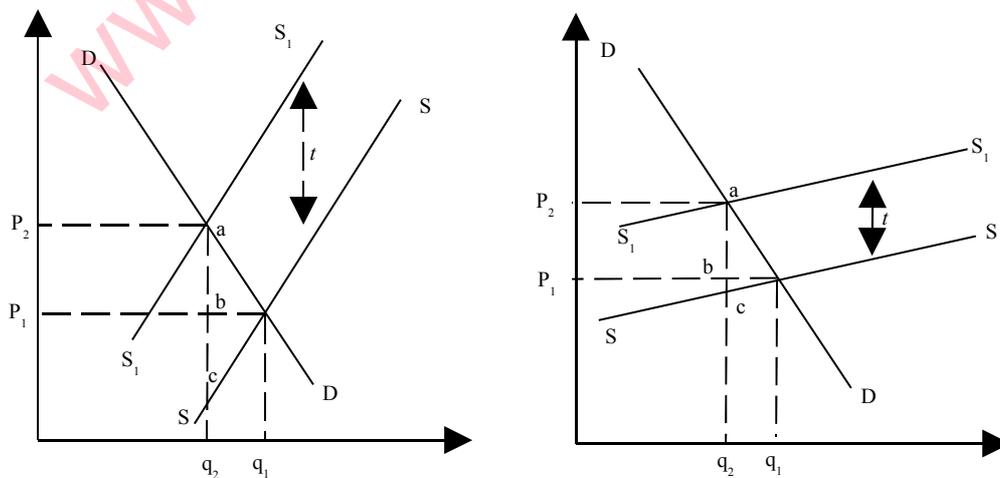
Consumers pay P_0P_1, AC of total tax revenue and producers pay P_2P_0CB .

The proportion of the tax which is passed on to the consumer rather than being borne by the supplier depends upon the elasticities of demand and supply in the market.

Figures 5(a) and 5(b) below illustrate the extreme cases of perfectly elastic demand and perfectly inelastic demand respectively.



The elasticity of supply is also relevant. Figure 6 shows that for a given demand curve, the more inelastic is the supply curve, the greater is the proportion of the tax that is borne by the supplier. Figure 6(a) shows a relatively inelastic supply curve SS. The tax per unit of t shifts the supply curve vertically upwards to S_1S_1 and the equilibrium price rises from p_1 to p_2 . The price to the consumer rises by ab per unit, while the supply price to the producer falls by bc per unit. Thus, the greater burden is borne by the supplier. Figure 6(b) in contrast shows a relatively elastic supply curve S_1S_1 , with a tax per unit of t , the same amount of tax per unit as in Figure 6(a). The supply curve shifts to S_1S_1 and the equilibrium price rises to p_2 . The price to the consumer rises by ab per unit, and the supply price to the producer falls by bc per unit. It can be seen from Figure 6 that the consumer bears a greater proportion of the tax burden the more elastic is the supply curve. Figure 6 also shows that, for a given demand curve, the price rise and the fall in the equilibrium quantity will both be greater when the supply curve is more elastic.



In general, the greater the elasticities of demand and supply, the greater will be the effect of a tax in reducing the quantity sold in and therefore produced for the market,

It can be shown that:

$$\frac{\text{consumers' share of tax}}{\text{producers' share of tax}} = \frac{\text{elasticity of supply}}{\text{elasticity of demand}}$$

(A proof of this is beyond the scope of this book -though not beyond the powers of a mathematically nimble student. You may care to try it as an additional exercise.)

Thus if a selective indirect tax of 10p is placed on a product where supply elasticity is 1.2 and demand elasticity is 0.8, the price of the good would rise by 6p, i.e.:

- (a) the consumer would pay 6p of the 10p (ratio 1.2: 0.8);
- (b) the supplier would pay 4p because the price increase of 6p is insufficient to pay the full tax of 10p per unit.

Further points to note are that:

- (a) since such a tax reduces output, it may be harmful to an industry. For some companies, the reduction in quantities produced may lead to significant rises in the unit costs of production. This could have adverse consequences on the competitive position of the firm if it competes in domestic or overseas markets with foreign firms which are not subject to the same tax;
- (b) indirect taxation may be used to create an improvement in the allocation of resources when there are damaging externalities.

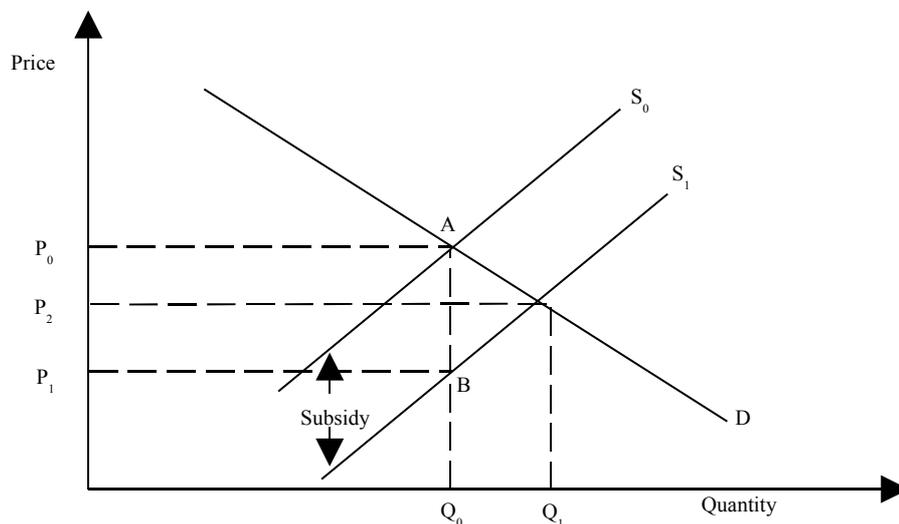
Subsidies

A subsidy is a payment to the supplier of a good by the government. The payment is made:

- (a) to encourage more production of the good, by offering a further incentive to suppliers; or
- (b) to keep prices on the market lower. Subsidised goods are cheaper to the consumer. Subsidised goods are therefore socially desirable goods whose production the government wishes to encourage; or
- (c) to protect a vital industry such as agriculture, when current demand in the short term is low and threatening to cause an excessive contraction of the industry.

A subsidy is rather like indirect taxation in reverse. In Figure 7:

- (a) supply curve S_0 is what the supply curve would be if no subsidy existed;
- (b) supply curve S_1 to the right of S_0 , is the supply curve with the subsidy.



If there was no subsidy, the free market equilibrium price would be P_0 , and output Q_0

With a subsidy per unit equivalent to AB , suppliers would be willing to produce Q_0 . If output were Q_0 then consumers need only be charged P_1 , because the supplier will receive a subsidy of AB per unit produced. This reflects the fact that the supply curve shifts from S to S_1 . But, given the position of the demand curve, there will be a shift in the equilibrium quantity produced to Q_2 , which can be sold on the market for P_2 . Thus the effect of the subsidy will be to:

- (a) increase the amount supplied in equilibrium; and
- (b) decrease the price, but the decrease in price will be less than the value of the subsidy itself.

Subsidies can be used to create an improvement in the allocation of resources when there are beneficial externalities.

When there is a beneficial externality, with the social benefits from supply of a good exceeding the private benefits, the 'ideal' equilibrium price is at:

- (a) a higher output; and
- (b) a lower price;

than the free market equilibrium price and output.

The effect of a subsidy is to shift the market equilibrium to:

- (a) a higher output; and
- (b) a lower price.

If the amount of the subsidy were made equal to the size of the beneficial externality, the effect of the subsidy would be to regulate the market in such a way that an ideal allocation of resources to the good is achieved. When the state does this, it might be said to be 'internalising the externality'.

Theory Of The Firm: Production And Costs

Your objectives

After completing this chapter you should:

- (a) understand the concepts of total costs, average costs and marginal costs and be aware of their effects on short-run decision-making;
- (b) understand the concept of economies of scale and how it affects decision-making in the long run.

Costs of production and the firm

Introduction

In this chapter we shall look at the costs and output decisions of an individual firm. In other words, we shall look at what the costs of production are for a single firm, and how these are affected by both short-run and long-run factors.

We shall then go on to consider profit and how much output a firm will produce at a given market price. The aggregate amount of goods supplied by every individual firm adds up to the market supply; by studying an individual firm we are looking at the 'building blocks' of market supply.

The firm is a wide term for any organisation which carries on business. In spite of their structural differences, firms will be treated as single, consistent decision-taking units, and we shall ignore any differences in decision-making procedures and economic structures between them. In particular, we will assume that the key objective of a firm is to maximise its profits.

Production is carried out using the factors of production which must be paid or rewarded for their use. The cost of production is the cost of the factors that are used.

<i>Factor of production</i>	<i>It's cost</i>
Land	rent
Labour	wages
Capital	interest
Entrepreneurship	normal or 'pure' profit

Fixed and variable inputs and costs

Firms combine various input resources to produce a given level of output. By varying the amounts of input used, the level of output can be altered. However not all inputs are equally flexible. Energy, raw materials, the number of labour hours and so on can be combined with each other with a great deal of flexibility, but the size of the factory or number of machines cannot be verified so quickly.

Economist take the view that in the short run some factors of production (or production inputs) are variable in supply and so have variable costs. Typically labour is regarded as a variable cost item. On the other hand some factors of production are fixed in supply and so have fixed cost item.

More precisely, fixed costs are those costs which do not vary directly with output, but which remain constant whether anything is produced or not. Variable costs are those which do vary directly with the level of output.

In the long run however the supply of short term fixed cost items can be made to vary and so all factors of production are variable. More land and buildings can be obtained, more capital obtained and more entrepreneurship stimulated.

In the short run, profit maximising decisions must be taken within the restriction of having some resources in fixed supply. In the longer run however most cases are variable because the supply of skilled labour, machinery, buildings and so on can be increased or decreased. Profit-maximising decisions in the long term are the fore subject to fewer restrictions about resource availability.

Short-run costs: total costs, average costs and marginal costs

Let us now turn our attention to short-run costs -i.e. costs of output during a time period in which only some resources of production are variable in availability and some resources of production are fixed in quantity.

There are three aspects of cost to be considered.

- (a) Total cost -TC.
- (b) Average cost -AC. Average cost is simply the total cost divided by the total quantity produced
 - (i) Average cost is made up of an average fixed cost per unit plus an average variable cost per unit.

$$AC = \frac{TC}{n} = \frac{TFC}{n} + \frac{TVC}{n}$$

(here n is the number of units produced)

In other words, $AC = AFC + A VC$

- (ii) Average fixed cost per unit (AFC) will get smaller as more units are produced. This is because TFC is the same amount regardless of the volume of output, so as n gets bigger, AFC must get smaller.
- (iii) Average variable costs per unit (A VC) will change as output volume increases.
- (c) Marginal cost -MC. This is the extra cost of producing one more unit of output. For example, the marginal cost for a firm of producing the 50th unit of output is the extra cost of making the 50th unit, having already made the previous 49 units. In other words the MC of the 50th unit is the total cost of making the first 50 units minus the total cost of making the first 49 units.

Numerical illustration of TC, AC and MC

Let us suppose that a firm employs a given amount of capital which is a fixed (invariable) input in the short run: in other words, it is not possible to obtain extra amounts of capital quickly. The firm may combine with this capital different amounts of labour, which we assume to be an input which is variable in the short term. Thus fixed capital and variable labour can be combined to produce different levels of output.

Here is an illustration of the relationship between the different definitions of the firm's costs: (the figures used are hypothetical).

Table 1

Units of output n	Total cost TC £	Average cost AC £	Marginal cost MC £
1	1.10	1.10	1.10
2	1.60	0.80	0.50 (1.60 - 1.10)
3	1.75	0.58	0.15 (1.75 - 1.60)
4	2.00	0.50	0.25 (2.00 - 1.75)
5	2.50	0.50	0.50 (2.50 - 2.00)
6	3.12	0.52	0.62 (3.12 - 2.50)
7	3.99	0.57	0.87 (3.99 - 3.12)
8	5.12	0.64	1.13 (5.12 - 3.99)
9	6.30	0.70	1.18 (6.30 - 5.12)
10	8.00	0.80	1.70 (8.00 - 6.30)

- (a) Total cost (TC) is the sum of labour costs plus capital costs, since these are by assumption the only two inputs.
- (b) Average cost (AC) is the cost per unit of output, i.e. $AC = \frac{TC}{\text{output}} = \frac{TC}{n}$
- (c) Marginal cost (MC) is the total cost of producing n units minus the total cost of producing one less unit, i.e. (n-1) units.

Note the following points on this set of figures.

- (a) Total cost. Total costs of production carry on rising as more and more units are produced.
- (b) Average cost. AC changes as output increases. It starts by falling, reaches a lowest level, and then starts rising again.
- (c) Marginal cost. The MC of each extra unit of output also changes with each unit produced. It too starts by falling, fairly quickly reaches a lowest level, and then starts rising.
- (d) AC and MC compared. At lowest levels of output, MC is less than AC. At highest levels of output, though, MC is higher than AC. There is a 'cross-over' point, where MC is exactly equal to AC. In this small example, it is at 5 units of output.

Economists' and accountants' concepts of cost

To an economist, cost includes an amount for normal profit which is the reward for entrepreneurship. To an accountant, there is no profit element within cost.

Another feature of cost accounting is that costs can be divided into fixed costs and variable or marginal costs. Total fixed costs per period are a given amount, regardless of the volume of production and sales. The variable cost per unit is a constant amount, so that the total variable cost of sales is directly proportional to the volume of sales.

Economists do not take this view. In the short run, there are fixed costs and variable or marginal costs, but the marginal cost of making each extra unit of output need not be the same for each extra unit that is made. In other words, the marginal cost per unit is not a 'standard' value for every unit produced. The following points are worth stressing.

- (a) To the economist, cost includes an element for 'normal' profit.
- (b) To the economist, cost means opportunity cost. Normal profit, which is the cost of entrepreneurship, is an opportunity cost, because it is the amount of profit that an entrepreneur could earn elsewhere, and so it is the profit that he must earn to persuade him to keep on with his investment in his current enterprise.
- (c) The short-run marginal cost per unit can change as more output is produced.

Average costs, marginal costs and diminishing returns

The relationship between AC and MC

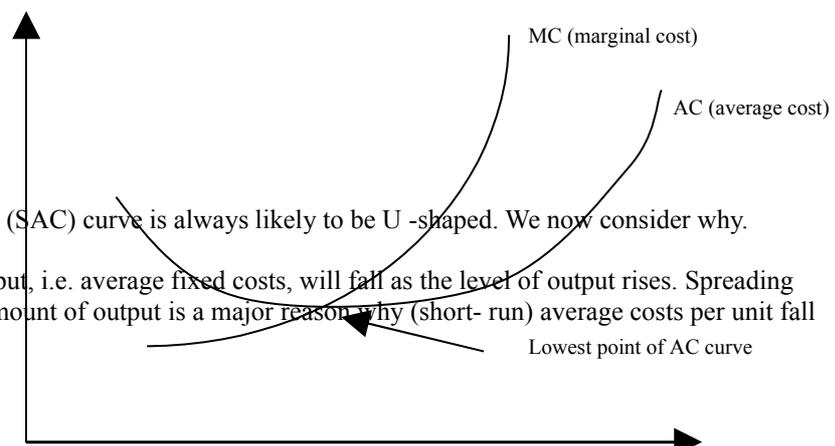
The relationships between average and marginal costs are important.

- (a) When the average cost schedule is rising, the marginal cost will always be higher than the average cost.

This makes sense. If the marginal cost of making one extra unit of output exceeds the average cost of making all the previous units, then making the extra unit will clearly cause an increase in the average unit cost.

- (b) When the average cost curve is falling, the marginal cost lies below it.
- (c) When the average cost curve is horizontal, marginal cost is equal to it.

The marginal cost curve always cuts through the average cost curve at the lowest point of the average cost curve (Figure 1).



The short-run average cost (SAC) curve is always likely to be U-shaped. We now consider why.

Fixed costs per unit of output, i.e. average fixed costs, will fall as the level of output rises. Spreading fixed costs over a larger amount of output is a major reason why (short-run) average costs per unit fall as output increases.

Variable costs are made up from the cost of the factors of production whose use can be varied in the short-run (for example wages, fuel bills and raw material purchases). Total variable costs therefore vary with output in the short run as well as in the long run.

- (a) The standard assumption about short-run variable costs is that up to a certain level of output, the variable cost per unit is more or less constant (eg wages costs and materials costs per unit of output are unchanged). If the average fixed cost per unit is falling as output rises and the average variable cost per unit is constant, it follows that the average total cost per unit will be falling too as output increases.
- (b) Nevertheless, there is evidence that average variable costs rise when output increases beyond a normal capacity level. Beyond this level greater output can be attained only by using larger quantities of the variable factors of production, but this will lead to problems such as overcrowding, managerial problems, and more frequent breakdowns of machinery from intensive use. Average variable costs will therefore begin to rise at some point, even assuming that there are no overtime payments or use of more skilled labour. With average variable costs per unit rising as output increases, the average total cost per unit will rise too.

Combining fixed and variable costs gives us the normal U-shaped short-run average cost curve.

The law of diminishing returns

Diminishing returns explain why a short-run average cost curve begins to rise at a certain level, and the average cost per unit of production gets higher as more output is produced.

The law of diminishing returns states that, given the present state of technology, as more units of a variable input factor are added to input factors that are fixed in supply in the short run, the resulting increments to total production will eventually and progressively decline.

In other words, as more units of a variable factor are added to a Quantity of fixed factors, there may be some increasing returns or constant returns as more units of the variable factor are added, but eventually, diminishing returns will set in.

The law of diminishing returns can also be expressed as the law of variable proportions, which states that as the proportions of a variable input factor to a fixed input factor are altered, and more of the variable factor is added to the fixed factor, the marginal product attributable to each extra unit of the variable factor will increase at first, but will later diminish and may eventually become negative.

Two important points to note about the law of diminishing returns are as follows.

- (a) It relates to the short-run situation, when some inputs are in fixed supply. It does not relate to the long run.
- (b) It is not a law that can be proved, but it has been found to apply frequently in practice

Marginal productivity

Diminishing returns can be explained by the marginal productivity of extra Quantities of a variable factor of production.

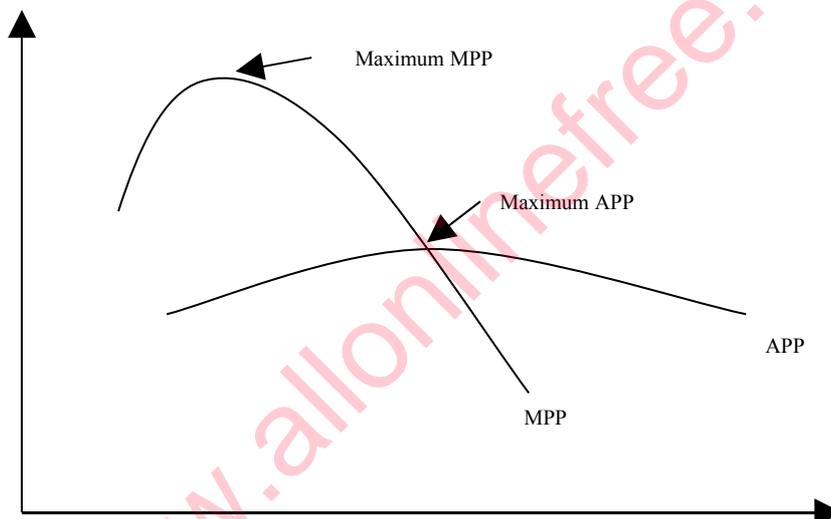
Total physical product is the total output, measured by quantity rather than value, that is produced by a given quantity of (fixed plus variable) production factors.

If we combine a fixed factor of production (capital, say, or land) with increasing quantities of a variable factor (labour, say) the total volume of output should increase as more and more units of the variable factor are employed. In other words, total physical product will increase as extra units of the variable factor are employed.

The average physical product (APP) of the variable factor of production will be:

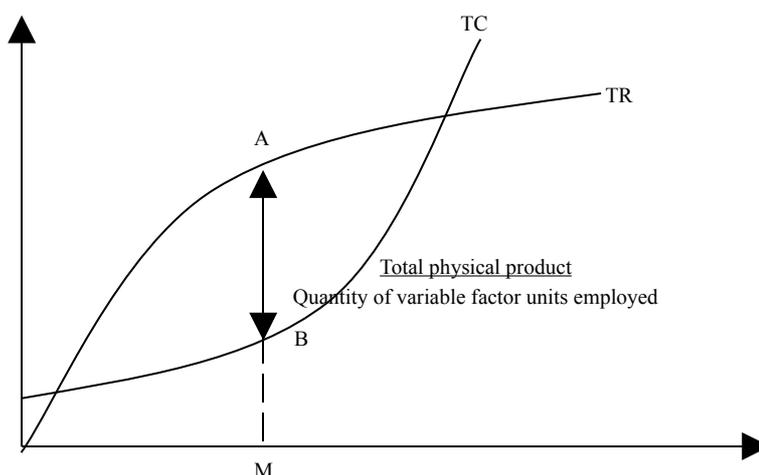
In the short-run, to produce more output, a firm must acquire extra quantities of variable factors of production, but fixed factors of production cannot be increased. Variable factors will be combined with fixed factors to produce the firm's output.

- (a) The extra physical output produced by an extra unit of variable factor is called the marginal physical product (MPP). At low volumes of output, extra units of a variable factor might succeed in producing an increasing MPP.
- (b) Eventually extra units of variable factor will no longer be as productive and MPP will start to get smaller. Returns from the quantities of variable factor will therefore get smaller or diminish – hence the name ‘law of diminishing returns’
- (c) A declining MPP (and a declining average physical product APP) explains why marginal costs of extra units of production will eventually increase and why average costs will start to increase beyond a certain level of output.
- (d) This is why the law of diminishing returns explains the U shape of the short run AC curve.



Profit maximisation

We can define profits as total revenue minus total economic costs at any level of output. Profits are at a maximum where the (vertical) distance between the total revenue (TR) and total cost (TC) curves in Figure 3 is greatest



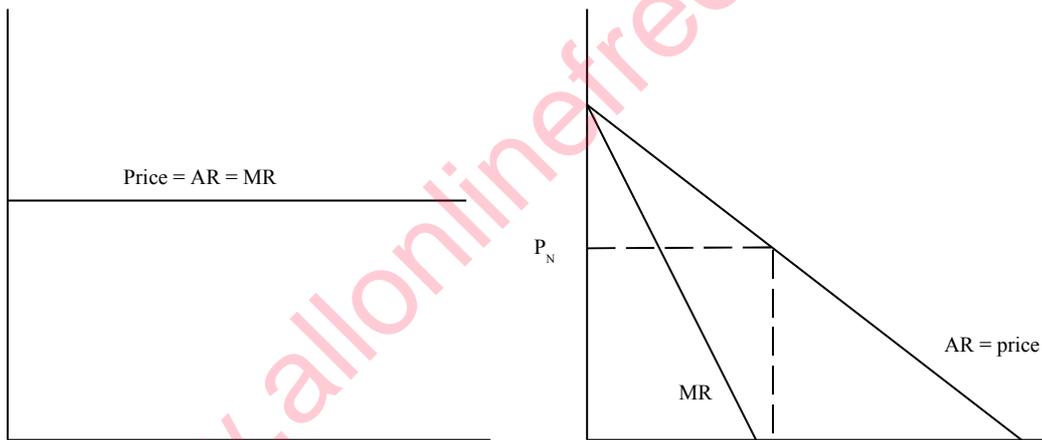
Total revenue, average revenue and marginal revenue

Before looking at how a firm decides its profit-maximising output, we must look at its revenues. There are three aspects of revenue to consider .

- (a) Total revenue (TR) which is the total income obtained from selling a given Quantity of output. We can think of this as Quantity sold multiplied by the price per unit: $TR = P \times Q$.
- (b) Average revenue (AR) which we can think of as the price per unit sold: $AR = \frac{TR}{n}$
- (c) Marginal revenue (MR) which is the addition to total revenue earned from the sale of one extra unit of output: $MR = p = AR$.

When a firm can sell all its extra output at the same price, the AR 'curve' will be a straight line on a graph, horizontal to the x axis. The marginal revenue per unit from selling extra units at a fixed price must be the same as the average price. (See Figure 4).

When the AR is falling as more units are sold, the MR must be less than the AR. In other words, if the price per unit must be lowered to sell more units, then the marginal revenue per unit obtained from selling the extra units will be less than the previous price per unit. (See Figure 5).



Note. In Figure 5, all units are sold at the same price. The firm has to reduce its price to sell more, but the price must be reduced for all units sold, not just for the extra units. This is because we are assuming that all output is produced for a single market, where a single price will prevail.

Illustration: AR and MR

This simple illustration shows how MR falls as AR falls, and MR is less than AR when this happens. .

<i>Units of sale</i>	<i>Price per unit</i>	<i>Total revenue</i>	<i>Marginal revenue</i>
Q	AR	(AR x Q)	MR
	£	£	£
1	7.0	7	7
2	6.5	13	6
3	6.0	18	5
4	5.5	22	4
5	5.0	25	3

Marginal revenue and price elasticity of demand

Marginal revenue and price elasticity of demand can be related mathematically through the following formula.

$$MR = P \left[1 - \frac{1}{\epsilon} \right]$$

where MR is marginal revenue

P is price
 ϵ is the absolute value of price elasticity of demand (ie ignoring the minus sign).

The formula is useful because it shows that

- (a) when $\epsilon = MR=P$
- (b) when $\epsilon = MR=0$
- (c) when $\epsilon > MR$ is positive
- (d) when $\epsilon < MR$ is negative.

When the price per unit has to be reduced in order to increase the firm's sales the marginal revenue can become negative. This happens in Figure 5 at price P_N when a reduction in price does not increase output sufficiently to earn the same total revenue as before. In this situation, the price elasticity of demand would be inelastic.

Marginal analysis

Marginal decision-making is based on the principle of measuring the change in benefits and/or costs that would arise as a consequence of a decision. The change would be worth having if the extra benefits exceeded the extra disadvantages.

Marginal analysis is based on some fairly simple rules and arithmetical relationships.

Rule 1. When there are no resource limitations, the scale of an activity should be increased if the marginal benefits from the change or increase exceed the marginal costs. Essentially, any activity is worth undertaking if it adds more benefits than it does cost.

In the case of a firm, for example, it would be worth spending extra money on advertising, given no restrictions on the cash available to spend on advertising, provided that the extra profits from the resulting extra sales* were higher than the extra cost of the advertising.

(* i.e. extra profits before deducting advertising costs),

Rule 2. Given no resource constraints, it follows that the scale or amount of an activity should be increased if the marginal net yield is positive.

Rule 3. When there is a shortage of resources, the scale of activity will be restricted to a level where marginal benefits still exceed marginal costs. In this situation, the decision problem will often be to choose between alternative competing courses of action, ie between mutually exclusive options.

In the case of a firm with limited resources, should the resources be put into making and selling product A or product B or product C? For example, if a firm has £1 to spend, and can spend it on advertising to yield extra profit of £2, or on extra staff to yield extra profits of £1.50, the choice would obviously be to spend the money on advertising.

To achieve optimisation, each activity should be taken to the level where the marginal (net) yield is the same for every activity.

A decision in favour of a course of action should be taken if the incremental revenues arising from the chosen action exceed the incremental opportunity costs. This is an application of marginal analysis.

Profit maximisation: $MC = MR$

As a firm produces and sells more units, its total costs will increase and its total revenues will also increase (unless the price elasticity of demand is inelastic and the firm faces a downward sloping AR curve).

- (a) Provided that the extra cost of making an extra unit is less than the extra revenue obtained from selling it, the firm will increase its profits by making and selling the extra unit.
- (b) If the extra cost of making an extra unit of output exceeds the extra revenue obtainable from selling it, the firm's profits would be reduced by making and selling the extra unit.
- (c) If the extra cost of making an extra unit of output is exactly equal to the extra revenue obtainable from selling it, bearing in mind that economic cost includes an amount for normal profit, it will be worth the firm's while to make and sell the extra unit. And since the extra cost of yet another unit would be higher (the law of diminishing returns applies) whereas extra revenue per unit from selling extra units is never higher, the profit-maximising output is reached at this point where $MC = MR$.

In other words, given the objective of profit-maximisation:

- (a) If MC is less than MR profits will be increased by making and selling more.
- (b) If MC is greater than MR profits will fall if more units are made and sold, and a profit-maximising firm would not make the extra output.
- (c) If $MC = MR$, the profit-maximising output has been reached, and this is the output Quantity that a profit-maximising firm will decide to supply.

Economies of scale and long-run costs

Introduction

We have not yet considered a firm's long-run costs of output. In the long-run, all inputs are variable, so the problems associated with the diminishing returns to variable factors do not arise; in other words, the law of diminishing returns applies only to short-run costs and not to long-run costs. Whereas short-run output decisions are concerned with diminishing returns to scale given fixed factors of production, long-run output decisions are concerned with economies of scale when all factor inputs are variable.

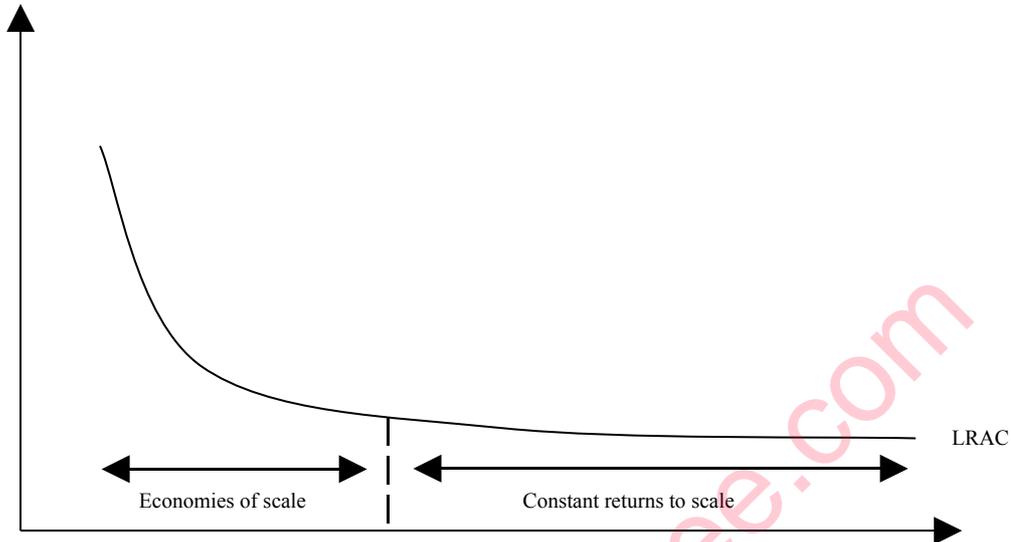
Output will vary with variations in inputs, such as labour and capital.

- (a) If output increases in the same proportion as inputs (e.g. doubling all inputs doubles output) there are constant returns to scale.
- (b) If output increases more than in proportion to inputs (e.g. doubling all inputs trebles output) there are economies of scale and in the long run average costs of production will continue to fall as output volume rises.
- (c) If output increases less than in proportion to inputs (e.g. trebling all inputs only doubles output) there are diseconomies of scale and in the long run average costs of production will rise as output volume rises.

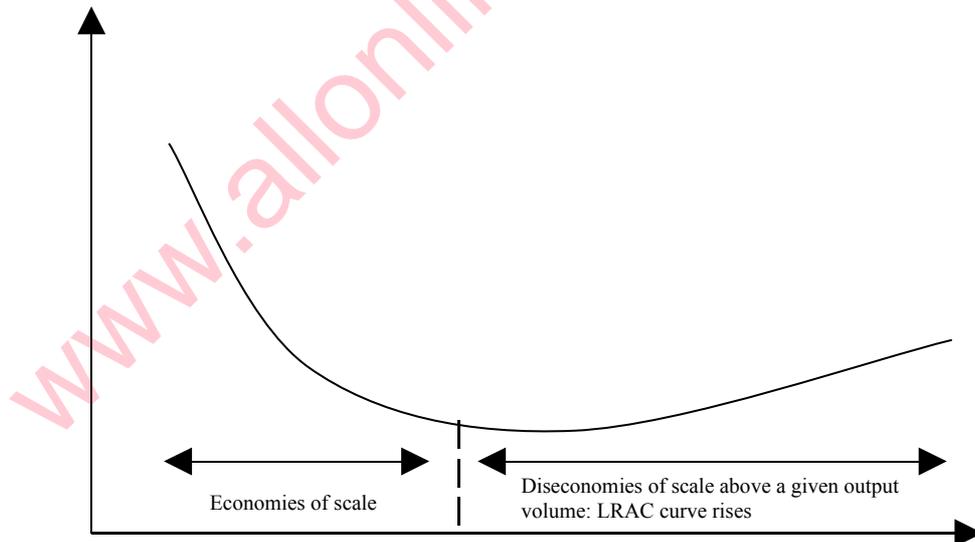
Returns to scale are, in effect, concerned with improvements or declines in productivity by increasing the scale of production (e.g. by mass-producing instead of producing in small batch quantities).

Economies of scale

Figure 6 shows the shape of the long-run average cost curve (LAC) if there are increasing returns to scale up to a certain output volume and then constant returns to scale thereafter. It may be that the flat part of the LAC curve is never reached, or it may be that diseconomies of scale are encountered.



Diseconomies of scale might arise when a firm gets so large that it cannot operate efficiently or it is too large to manage efficiently, so that average costs begin to rise.



A profit-maximising firm will try to minimise its average costs in the long-run, and to do this it will try to produce output on a scale where the LAC curve is at its lowest point. While there are economies of scale, a firm will always be trying to grow.

The economies of scale attainable from large-scale production may be categorised as:

- (a) internal economies of scale: economies arising within the firm from the organization of production; and
- (b) external economies of scale: economies attainable by the firm because of the growth of the industry as a whole.

Internal economies of scale arise from the more effective use of available resources, and from increased specialisation, when production capacity is enlarged.

- (a) Specialisation of labour: in a large undertaking, a highly skilled worker can be employed in a job which makes full use of his skills. In a smaller undertaking, individuals must do a variety of tasks, none of which they may do very well ('Jack-of-all-trades -master of none').
- (b) Division of labour: because there is specialisation of labour there is also division of labour, i.e. work is divided between several specialists, each of whom contributes his share to the final product. A building will be constructed, for example, by labourers, bricklayers, plumbers, electricians, plasterers, etc. Switching between tasks wastes time, and division of labour avoids this waste.
- (c) Large undertakings can make use of larger and more specialised machinery. If smaller undertakings tried to use similar machinery, the costs would be excessive because the machines would become obsolete before their physical life ends (i.e. their economic life would be shorter than their physical life). Obsolescence is caused by falling demand for the product made on the machine, or by the development of newer and better machines.
- (d) For a similar reason, large undertakings can use specialised tools which small undertakings would find too costly.

Economists refer to large capital items that are only economically justifiable at high volumes of output as indivisibles.

- (e) Dimensional economies of scale refer to the relationship between the volume of output and the size of equipment (e.g. storage tanks) needed to hold or process the output. The cost of a container for 10,000 gallons of product will be much less than ten times the cost of a container for just 1,000 gallons.
- (f) Buying economies may be available, reducing the cost of material purchases through bulk purchase discounts.
- (g) Indivisibility of operations: there are operations which:
 - (i) must be carried out at the same cost, regardless of whether the business is small or large; these are fixed costs and average fixed costs always decline as production increases;
 - (ii) vary a little~ but not proportionately, with size (i.e. having 'semi-fixed' costs);
 - (iii) are not worth considering below a certain level of output (e.g. advertising campaigns, marketing structures).

Set-up costs for batch production are an example of 'fixed cost' items for which average unit costs become lower as the size of the production run gets bigger .

- (h) Specialisation of labour and machines result in simplification and standardisation of operations (i.e. variety reduction) which themselves result in lower costs.
 - (i) Stock holding becomes more efficient. The most economic quantities of inventory to hold increase with the scale of operations, but at a lower proportionate rate of increase.

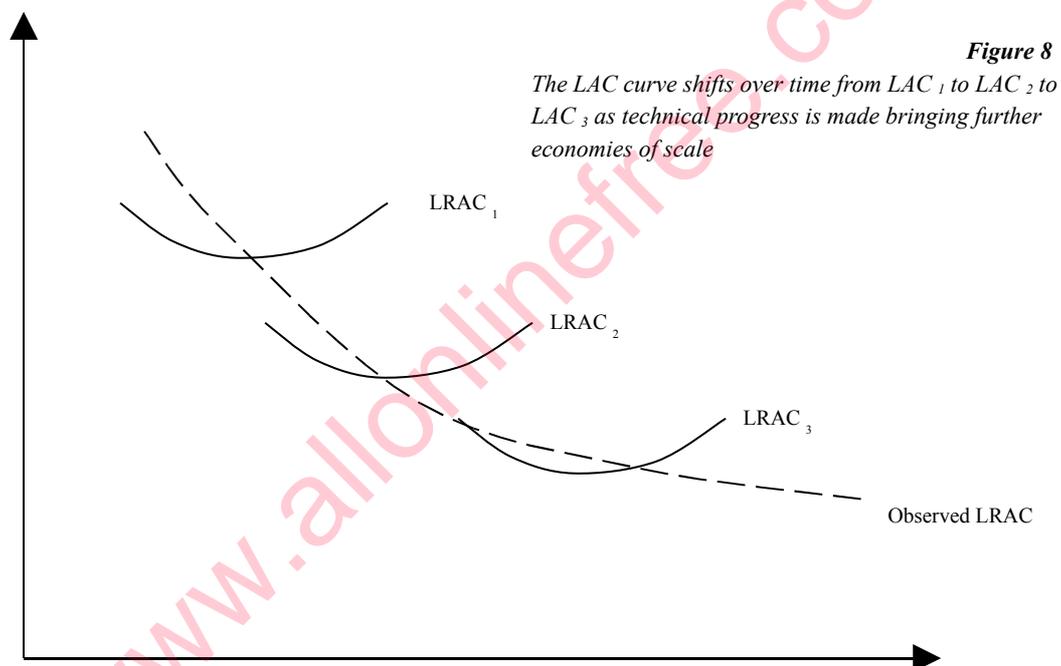
External economies of scale occur as an industry grows in size. For example:

- (a) a large skilled labour force is created and educational services can be geared towards training new entrants;
- (b) specialised ancillary industries will develop to provide components, transport finished goods, trade in by-products, provide special services etc. For instance, law firms may be set up to specialise in the affairs of the industry.

The extent to which both internal and external economies of scale can be achieved will vary from industry to industry, depending on the conditions with respect to that industry. In other words, big firms are better suited to some industries than others.

Technological progress and shifts in the long-run cost curve

The u-shaped LAC curve predicted by the economic theory of eventual diminishing returns to scale may not exist in fact because of technological progress. Technological progress would shift the LAC curves over time, as shown in Figure 8, so the LAC curve observed from empirical data over time would be the L-shaped curve shown by the dashed line.



The reasons why technological progress reduces long-run costs are that new technology reduces short-run costs, and since the long-run average cost curve represents a series of short-run cost curves at different output volumes, there will be reductions in the LAC too.

Short-run costs will fall because:

- (a) the new technology might help in achieving greater economies of scale; it will improve labour productivity too, measured as output per employee over time;
- (b) there might be a shift to using cheaper machines from relatively more expensive labour. In other words, new technology can result in lost jobs.

Factor Markets And Factor Rewards

Your objectives

After completing this chapter you should:

- (a) be aware of the factors of production and the market of the rewards accruing to each of them;
- (b) understand the application of supply and demand theory to factors of production.

Factors of production and their rewards

In the previous chapter we saw that the factors of production could be analysed into four categories: land, labour, capital and entrepreneurship. In this chapter we develop that analysis more fully.

Each scarce economic resource has a value, and the owner of the resource or factor of production is rewarded for giving it up to someone else. Firms are rewarded for the goods and services they produce by the price customers will pay for them. The resources used in production are also rewarded, by the price that firms pay for them.

- (a) Land is rewarded with rent. Although it is easy to think of land as property, the economic definition of land is not quite what you might suppose. Land consists not only of property (the land element only: buildings are capital) but also the natural resources that grow on the land or that are extracted from the land (i.e. the natural resources of the soil and woodlands and extracted minerals such as coal).
- (b) Labour is rewarded with wages. Labour consists of both the mental and the physical resources of human beings. (Salaries as well as weekly-paid wages are defined collectively as 'wages' in economics.)
- (c) Capital is rewarded with interest. It is easy to think of capital as financial resources, and the rate of interest is the price mechanism in balancing the supply and demand for money. However, capital in an economic sense is not 'money in the bank'.

Capital refers to man-made items such as plant, machinery and tools which are made and used not for their own sake, but to aid the production of other goods and services. The cost of using machinery and plant and so on is 'interest',

- (d) Entrepreneurship or enterprise is the fourth factor of production. An entrepreneur is someone who undertakes the task of organising the other three factors of production in a business enterprise, and in doing so, bears the risk of the venture. He creates new business ventures and the reward for the risk he takes is profit.

The cost of production (rent, wages, interest and profit) is the sum total of the rewards for all the factors of production that go into making a good.

Distribution theory

Distribution theory is concerned with how much reward each factor of production gets. What determines the amount that is paid for land, labour, capital and entrepreneurship? How is the total income 'cake' divided between them?

The total income earned by all factors of production within a national economy equals the national income. National income is the subject of a later chapter.

Factor prices

The prices paid to each factor of production are sometimes referred to as 'factor prices'. The prices for land, labour and capital are determined by supply and demand. Entrepreneurship and profit are different, and these will be discussed later.

Factor demand as a derived demand

The demand for factors of production is a derived demand.

By this we mean that the factors of production are not demanded for their own sake. They are demanded because a firm needs them to make goods, which are then sold to households. It is the demand by households for goods from which the demand by firms for the factors of production is derived.

Capital and interest

Interest is the reward for capital. Capital as a factor of production consists of:

- (a) stocks of finished goods;
- (b) producer goods (i.e. machines, tools, buildings, office equipment etc).

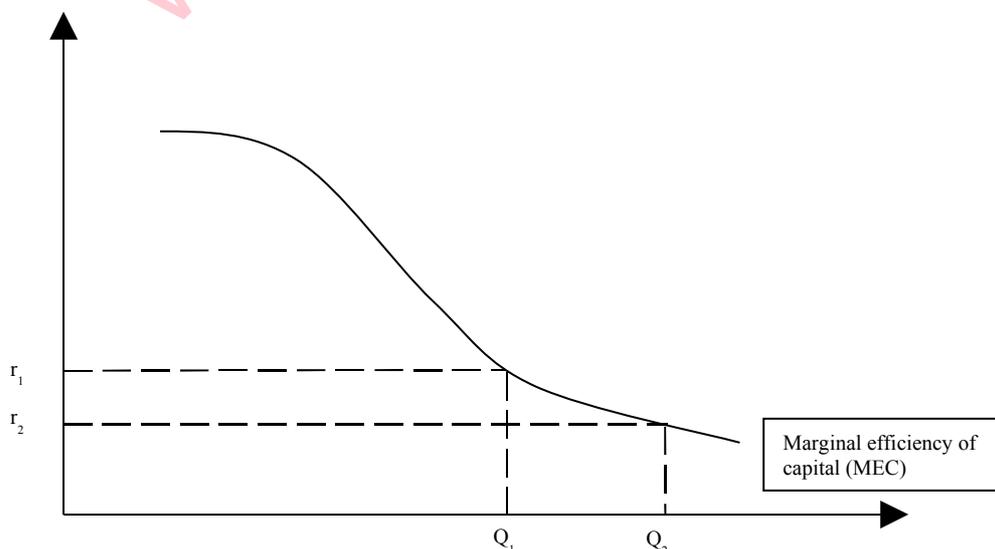
The rate of interest according to traditional economic theory is determined by supply and demand.

- (a) the demand for capital comes from firms which expect to invest in stocks and equipment so as to create more output, make more sales and earn more profit
- (b) the supply of capital (i.e. finance to acquire stocks and equipment etc) comes from investors

The marginal efficiency of capital

The demand by firms to borrow capital is explained in traditional economic theory by the marginal efficiency of capital. It is reasonable to assume that if firms borrow more and more capital to invest the additional investments that they make will become less and less profitable.

Firms should always seek to invest in the opportunities that offer the biggest returns and once these have been invested in remaining opportunities will not offer returns quite as big. As more investments are made the returns from additional capital investments will gradually decline. The marginal efficiency of capital refers to this declining size of return as the volume of investment increases.



The MEC curve for all the firms in an industry is the industry's demand curve for capital, and it is the sum of the demand curves (MEC curves) of all the individual firms in the industry. An industry's MEC curve and an individual firm's MEC curve have the same basic shape, as shown in Figure I.

The supply of capital

The rate of interest also depends on the supply of capital from investors, and the interaction of supply and demand establishes interest rates.

The supply of capital comes from savers. Savings are the resources that are needed to produce capital (i.e. to pay for the materials and labour that produce the capital) and which could have been used for current consumption instead. Savings lead to investment and the creation of capital, but savings are only made by 'sacrificing' some current consumption.

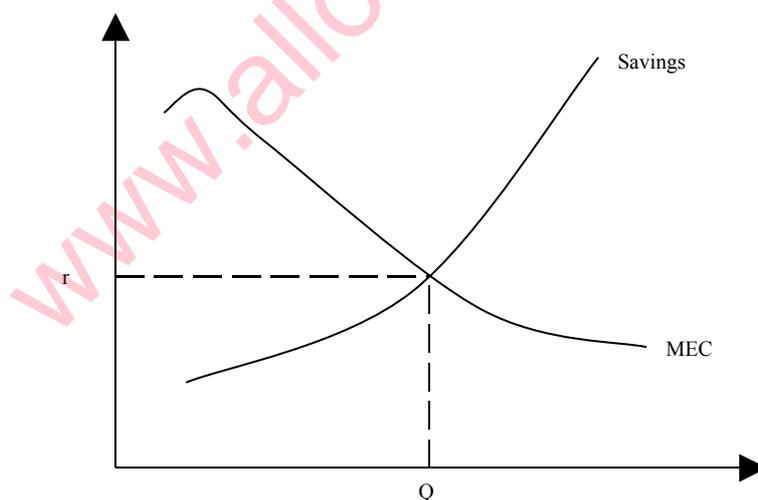
Savers choose to save in order to make possible the production of even more outputs in the future, and so the amount of savings is determined by comparing:

- (a) what the available wealth could be used to obtain now from the current consumption; and
- (b) how much extra wealth will be obtained in the future from saving.

This extra wealth in the future, which makes savers prefer to save rather than consume their wealth now, is represented by interest.

Higher interest will make saving more attractive, and the supply of savings will therefore increase. The price of capital (the interest rate) should therefore be determined by the interaction of supply (savings) and demand (marginal efficiency of capital).

In Figure 2, the equilibrium interest rate would be r with quantity Q of capital supplied by savers and demanded by firms.



Labour and wages

The demand for labour and marginal productivity theory

A similar demand and supply analysis can be made for labour and the price of labour (ie wages).

- (a) Like the demand for capital, the demand for labour by firms is a derived demand, arising from consumer demand for the firms' output.

- (b) Labour is employed to help to make the goods and services of the firm, and the more labour that is employed, the greater should be the total volume of goods or services produced.
- (c) However, with labour the law of diminishing returns applies, because as more and more labour is hired, the productivity of the extra workforce will gradually decline. A firm cannot keep on hiring extra labour, in the short term at least, and expect every additional recruit to contribute the same extra output to the firm as others in the workforce.

In our basic analysis of marginal productivity theory, we shall assume that the amount of other factors of production is in fixed supply, so that any additions to a firm's output and revenue can only come from additions to the labour force. In other words, labour is a variable factor of production.

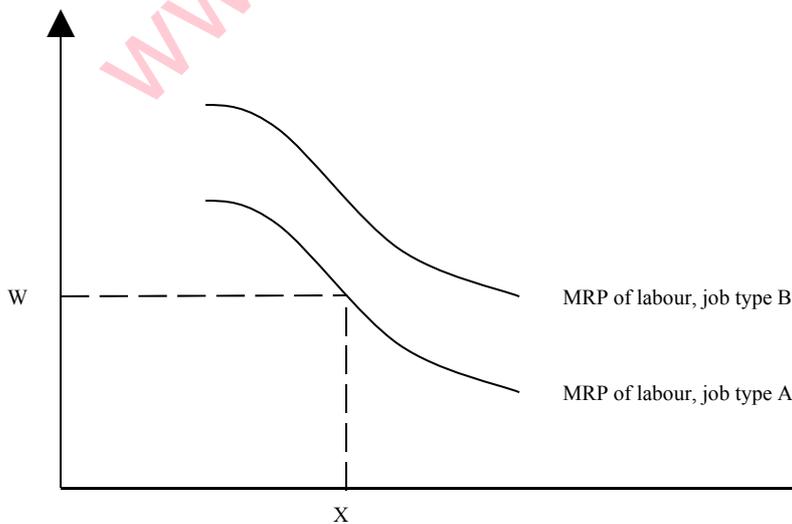
The marginal physical product (MPP) of labour is the additional units of output from one extra unit of labour. In accordance with the law of diminishing returns, the MPP of labour will decline as more and more output is produced.

The marginal revenue product (MRP) of labour is the marginal revenue value of the marginal physical product of labour. This is the extra revenue that firms in the industry would obtain from the extra output provided by each extra recruit to the workforce. Like the MPP, the MRP will be gradually declining. Table 1 below illustrates this relationship between MPP and MRP.

Table 1

Number of units of labour	Total output units	Revenue value of this output £	Marginal physical product	Marginal revenue product £
1	60	700	-	
2	110	1,200	50	500
3	150	1,500	40	300
4	180	1,700	30	200

The MRP of labour is similar in concept to the marginal efficiency curve of capital. It also represents the demand curve for labour by a firm (or by firms in the industry as a whole). Firms should be willing to pay for labour provided that the marginal revenue product of labour exceeds the cost of employing the labour. Quite simply, if a firm can make an extra £150 per week from hiring an extra employee, it should be willing to hire the employee provided that the weekly wage does not exceed £150. In Figure 3, if the MRP of labour for a certain type of job, job type A, is as shown, and the wage level for job type A is W, then the industry would want to employ X employees in job type A, because the MRP of labour exceeds the wage rate up to X.



In contrast, job type B has a higher marginal productivity value than job type A, and so the industry would be willing either:

- (a) to pay a higher wage for the same quantity of employees as job type A; or
- (b) to employ more employees into job type B than job type A if the wage rate for both were W.

This analysis of the MRP of labour only considers the demand by firms for labour of different types and skills. It does not consider the supply of labour.

According to marginal productivity theory, wage levels are determined by the interaction of the demand for and the supply of labour. The supply of labour is influenced by wage rates. Higher wages will attract more people willing to do the work.

The supply curve for labour can therefore be shown as the marginal cost of the labour, i.e. the extra total wage payments needed to increase total labour supply by each marginal extra amount.

For example, suppose that at a wage level of £160 per week, the supply of labour into a job would be 8 men. At a wage level of £170 per week, the willing supply would be 9 men, and at £180 per week, 10 men would be willing to do the job. The marginal cost of the 9th and 10th men would be as in Table 2.

Table 2

Wage = Average cost of labour (AC _L)	Supply (number of men)	Total wages	Marginal cost of labour (MC _L)
£		£	£
160	8	1280	-
170	9	1530	250 (1530-1280)
180	10	1800	270 (1800-1530)

Although the example is small, the figures do show that when higher wages must be paid to attract more labour:

- (a) the supply curve for labour, which is the marginal cost curve for labour (MCL), will be positive and therefore rising; and
- (b) MCL will be higher than the wage level (i.e. higher than the average cost of labour ACL).

Returning to marginal productivity theory, we therefore have wage levels determined by the interaction of supply and demand, which is where the MCL curve intersects with the marginal revenue product curve for labour (MRPL).

In a perfectly competitive industry, all extra quantities of labour can be obtained freely at a constant wage rate (for example, if a firm or an industry can hire limitless quantities of labour at a wage of £250 per week).

Figure 4
Competition among firms for labour
 Wage level = W, employment level in the industry = L_i
 and in an individual firm = L_f

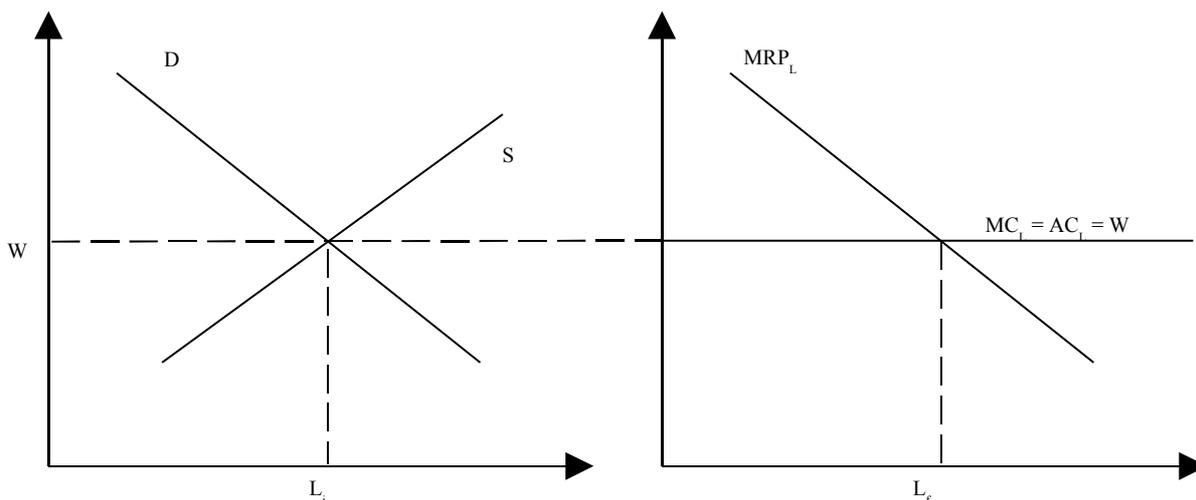
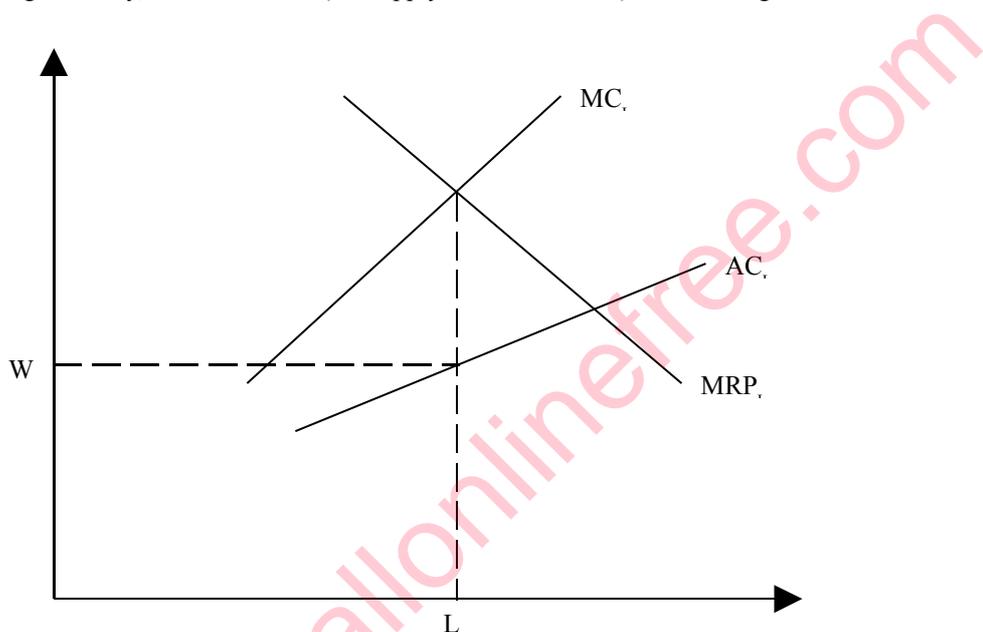


Figure 4 shows that an individual firm will continue to hire more labour at wage rate W until the MRP of labour falls to this level. The individual firm will therefore hire L_f units of labour. Total employment in the industry (L_i) is the sum of the employment of labour by the individual firms.

If the industry is dominated by a single firm, this firm is the only, or the main, 'buyer' of labour in the industry. Such a firm is said to be a 'monopsonist' buyer of labour.

MRPL

In this situation, the firm will pay lower wages if it needs fewer employees, and higher wages if it needs to attract more labour. The ACL curve rises with output/numbers employed, and more significantly, the MCL curve (the supply curve for labour) is also rising.



Demand and supply analysis shows that the number employed will be L , which is the level where $MCL = MRPL$.

Limitations of marginal productivity theory

The marginal productivity theory of wages cannot account wholly for the determination of wage rates and wage differentials because the assumptions on which it is based do not apply in reality.

- (a) It is often impossible to calculate the marginal productivity of labour, especially in administrative work or service industries e.g., bank clerks and shop assistants do not produce a measurable physical output.
- (b) Marginal productivity theory for labour assumes that all other factors of production are held in constant supply. This is unlikely to be so, especially in the case of capital. As the amount of labour employed changes, so too would the amount of capital.
- (c) A further assumption of the marginal productivity theory of wages is that labour is free to enter the market or leave it for alternative employment elsewhere. In practice, this might not be the case, and there might be imperfections in the labour market.

'Imperfections' in the labour market

'Imperfections' in the labour market prevent a free market in labour from operating, because of restrictions on the free supply of labour.

These labour supply restrictions include the following.

- (a) Society, laws and customs. (For example, the social custom might be for women to work, or not to work; or for children to work from a certain age. Some jobs might be more 'socially acceptable' than others.)
- (b) Labour immobility. (This refers to the movement of labour from one industry to another and from one geographical region or country to another.)
- (c) Barriers to entry into the trade or profession.
- (d) Lack of freely available information about jobs and wages,
- (e) Trade union influence on the supply of labour and wage rates.

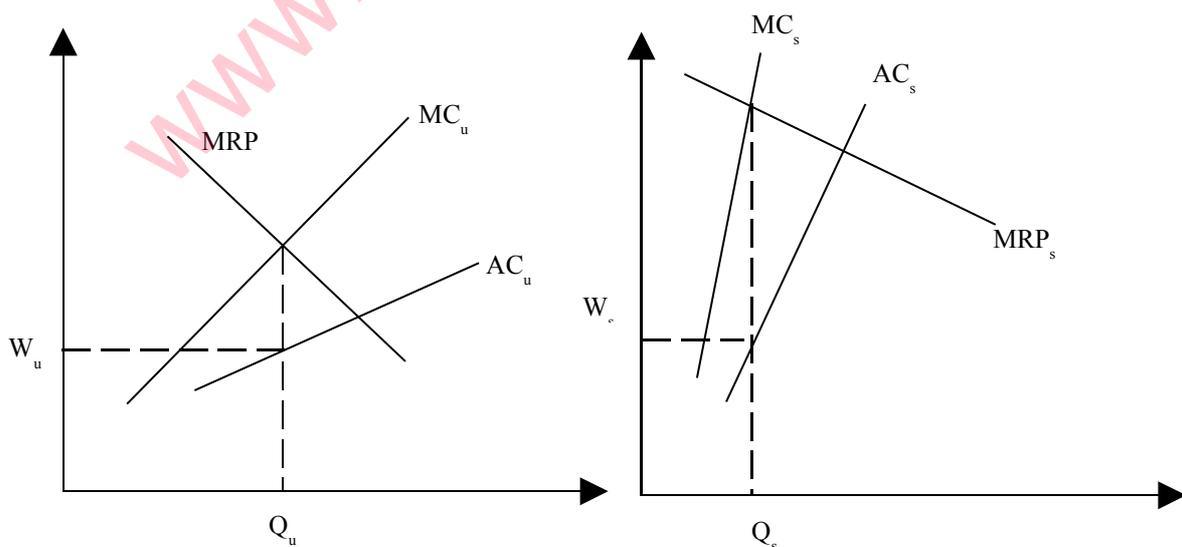
We shall now consider some of these in more detail

Wage differentials

Wage differentials are differences in the rate of pay between one type of job and another. Some jobs are more attractive than others. If wages were equal in all occupations the dirty and disagreeable jobs would attract few workers while most people would seek employment where conditions were pleasant. The tendency would therefore be for the wages in the disagreeable and dirty jobs to rise while wages in the more pleasant occupations would fall.

A theoretical explanation of wage differentials can be based on straightforward demand and supply analysis:

- (a) the demand for a particular type of worker comes from employees (or customers)
- (b) the supply of a particular type of worker comes from the individuals who are able and who choose to do that type of work.



Skilled workers are more productive and add more value to a firm's marginal revenue product than unskilled workers. In Figure 6, we assume that employers effectively form a buyers' cartel or 'monopsony'. The MRP of a skilled labour force is shown by MRP_s , and the MRP of an unskilled labour force is shown by MRP_u .

Skilled workers expect to be paid more for their skills. In order to attract a bigger supply of skilled workers, i.e. more people willing to acquire the necessary training, skills and qualifications, higher wages must be paid, and the supply curve of skilled labour (MC_s) will therefore lie to the left of the supply curve of unskilled labour (MC_u). The supply of skilled workers will also be more inelastic, because the barriers to entry (e.g. the need to obtain suitable training and qualifications) are higher, or the availability of individuals with suitable talent will be restricted. As a consequence, there will be a wage differential between the skilled and the unskilled workers, with skilled workers earning W_s and unskilled workers earning W_u .

How might wage differentials be eliminated? If a group of low-paid workers wished to eliminate the wage differentials between themselves and more highly paid skilled workers, they might try strike action, or even ask for government support (minimum pay legislation, or income controls). However, according to marginal productivity theory, the solution would be to increase productivity. If unskilled labour can become more productive, their MRP curve will shift to the right, and so they can justify higher wages. Wage differentials with other workers would be reduced and perhaps eliminated. This is why, in some highly profitable industries, unskilled workers might have such a high MRP that they earn higher wages than skilled workers in an unprofitable industry.

Marginal productivity theory is not the only way to explain wage differentials. Other factors affecting labour supply might help to influence wage levels for skilled workers - such as the willingness of some skilled labour to work for low wages, despite their skills and training (e.g. nurses and social workers).

Trade unions and the bargaining theory of wages

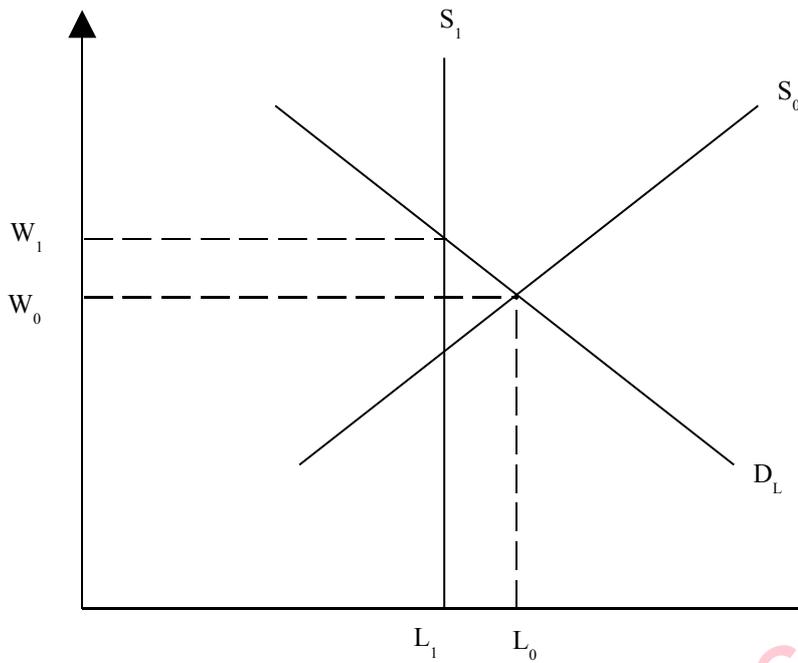
Although we have looked at marginal productivity theory and the view that wage levels are set by demand and supply factors, it should also be recognised that trade unions try to negotiate higher wages for labour. There is a 'bargaining theory' of wages which states that wage levels are set by negotiation between unions and management.

Collective bargaining is a term which refers to the process by which unions negotiate and reach agreements with employers. It is common for collective bargaining to involve a single monopolist seller of labour (the trade union) and one buyer or monopsonist (a single firm or an employers' federation). As such, annual wage claims may be one of a trial of strength between two 'giants'.

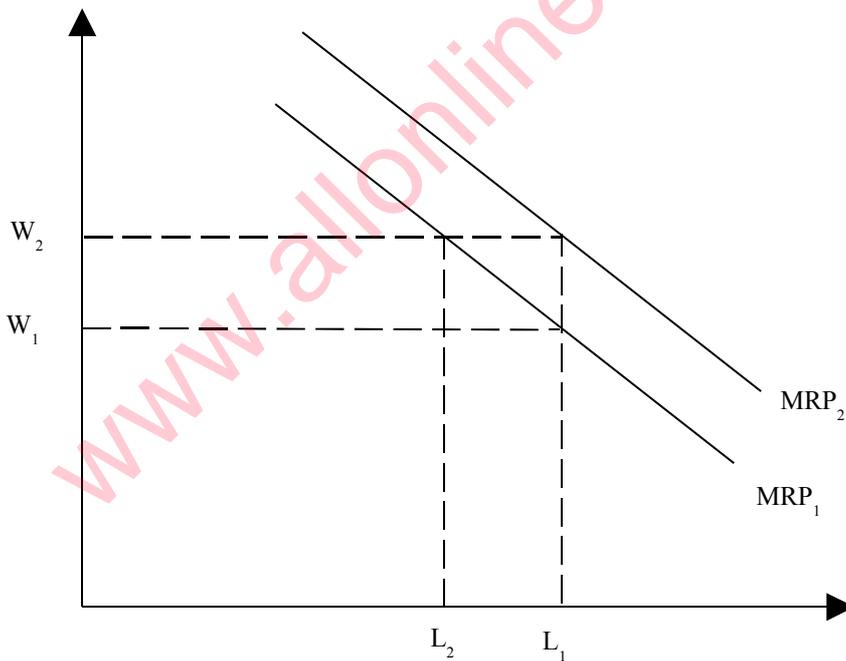
The role of trade unions, in economic terms, has two aspects:

- (a) to erect and maintain barriers to entry into jobs in the industry, thus ensuring high earnings for the existing members;
- (b) to monopolise the supply of labour in the industry. If the demand for labour is in the hands of a single employer or employers' federation, this can influence the price at which labour is bought. If the supply of labour is in the hands of a collective body as opposed to individuals, this can influence the price at which it can be sold.

By restricting entry to the labour force, trade unions can force wages to move from W_0 to W_1 , (see Figure 7) by effectively changing the supply curve from S_0 to S_1 . This will however result in fewer jobs. The number employed will fall from L_0 to L_1 .



However, once wage rates have been given an initial increase by the unionisation of the work force, any further pay rises, given no change in the marginal revenue product (MRP) of labour, will probably reduce the total demand for labour by employers.

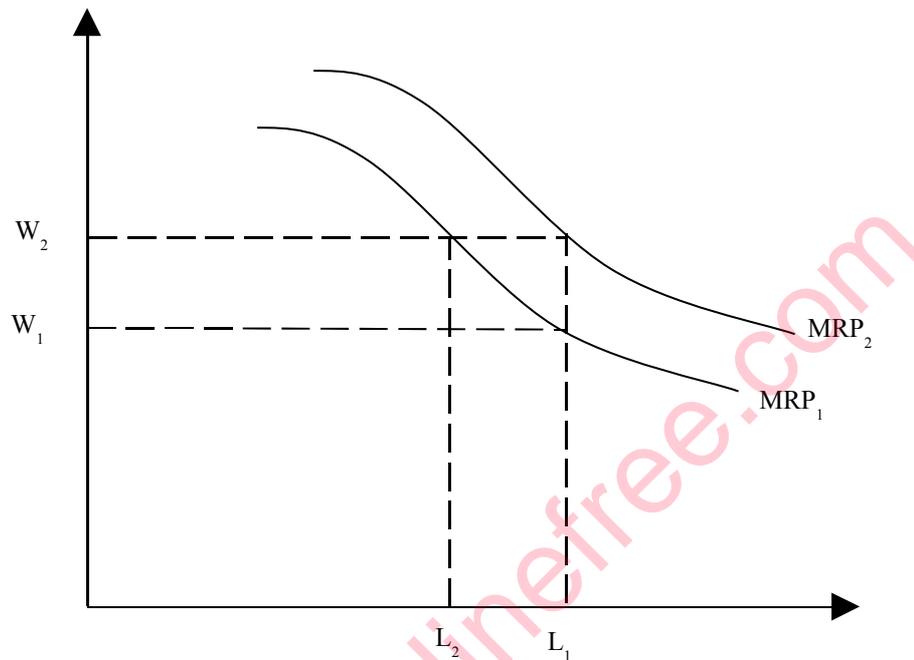


When wages rise from W_1 to W_2 , the demand for labour will fall from L_1 to L_2 (Figure 8), given no change in the MRP of labour (initially MRP_1). However, if the labour force agrees to an improvement in productivity so that the marginal revenue product of labour shifts to MRP_2 , an increase in wages from W_1 to W_2 could be achieved without changing the total workforce employed from L_1 .

Wage rises and inflation

A wage increase can have one of two effects

- (a) As illustrated in Figure 9 if the employer cannot raise the prices of his final product the marginal revenue product of labour cannot be increased and so the amount of labour demanded by the employer will fall from L_1 to L_2 .



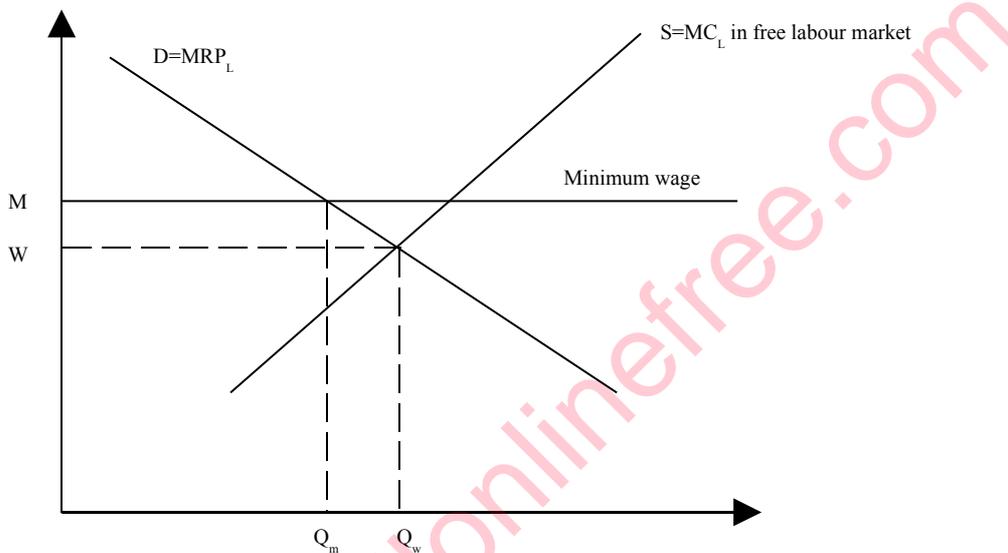
- (b) If the employer can pass the extra labour costs on the end customer by raising the price of the end product the MRP of the labour will be increased and so the same quantity of labour L_1 can be kept in employment at the new wage level as shown in figure 9

Minimum wages

In some industries of some countries, there is a minimum wage agreement which means that all workers in the industry must earn at least the minimum wage. Some trade unionists have argued for a national minimum wage. The purpose of a minimum wage is to ensure that low-paid workers earn at least enough in wages to have a certain standard of living.

If a minimum wage is enforced by legislation or negotiated nationally for an industry by a trade union the minimum wage will probably be above the current wage level for the jobs concerned. The consequences of a minimum wage would then be:

- (a) to raise wage levels for workers employed to a level above the 'equilibrium' wage rate; but
- (b) to reduce the demand for labour and so cause job losses.



Without a minimum wage, OQ_w workers would be employed at wage rate W (Figure 10).

Bargaining theory of wages: conclusion

It is probably fair to conclude that the bargaining theory of wages, that unions can fix a wage level by negotiating pay with employers, is not inconsistent with marginal productivity theory and the principles of demand and supply. The level of wages bargained by unions will help to determine employment levels (demand and supply) and when conditions in the industry 'favour' the employer with labour having low marginal revenue productivity, unions will find it difficult to bargain for higher wages or to preserve jobs in the industry.

Factors influencing demand for labour

As we have seen, demand for labour is influenced by the marginal productivity of labour. We now look at specific factors influencing demand and marginal productivity. These are:

- (a) the efficiency or productivity of labour;
- (b) the skills of labour;
- (c) the substitutability of capital for labour.

The efficiency of labour

Improvements in productivity will increase the demand for labour, and productivity in turn depends on:

- (a) the attitudes of the individual;

- (b) the intelligence and skills of individuals;
- (c) the attitudes of the work group to which the individual belongs;
- (d) established work practices and trade union influence;
- (e) the skill of management in getting the best out of a workforce;
- (f) effective use of other factors of production;
- (g) possibly, greater specialisation (i.e. an increasing 'division of labour').

Labour skills: specialisation of labour

Specialisation of labour helps to increase labour productivity. Work is divided up into small units, and specialist labour concentrates on a particular unit of the overall work. Plant and equipment can be specialised too, and automation helps to increase labour productivity. Work can be organised effectively by assigning specialists to each aspect of what has to be done, and time is saved because workers do not have to keep switching from one job to another.

The substitutability of capital for labour

Substitution between factors of production (e.g. between labour and capital) will take place:

- (a) provided that substitution is practical or technologically feasible (e.g. that machines can be made to do the work previously carried out by labour, or that labour can physically do the work of machines);
- (b) when the price or productivity of one factor of production rises relative to another. If wages go up, the marginal cost of labour will rise, and firms will want less labour at this higher cost. Labour will also become more expensive in relation to capital, and there will be some substitution of capital for labour. The net result of an increase in wages will be a reduction in the Quantity of labour employed -unless the productivity of labour can be increased at the same time, to strengthen the demand for labour.

The elasticity of demand for labour

The change in demand for labour in response to a change in wage rates can be measured by the elasticity of demand for labour. This is:

$$\frac{\% \text{ change in numbers employed}}{\% \text{ change in wages}}$$

The factors influencing the degree of elasticity of demand for labour are:

- (a) the technical ease with which employers could substitute other factors of production (mainly capital) for labour;
- (b) the elasticity of demand for the final product. If the product made by the work force has an inelastic demand, producers can pass on higher wage costs more easily to consumers by raising prices. However, if demand for the end-product has a high price elasticity, an increase in wages will result, through higher prices, in a sharp fall in demand for labour;
- (c) the elasticity of supply of alternative factors of production. Even if it is technically possible to substitute labour with, say, capital, it might be too costly for producers to do so if the elasticity of supply of capital is low. An increase in supply would then only be achieved by paying significantly more for the substitute factor (capital), and so in spite of higher wages costs, it might still be less costly to use labour than to switch to capital as a substitute factor;
- (d) the proportion of labour costs to total costs.

Land and rent

Rent

The price of land, which is rent, is also determined by supply and demand. In discussing land, it is important to get our basic definitions clear. In everyday language, if a person buys some land, we probably mean that he buys some buildings with a supply of water, electricity and so on. To the economist, buildings and water supply are capital; land is the earth and its natural resources. It is also common to speak of renting a house, a car or a television. This is 'commercial rent' paid to the landlord who is an owner of capital. Commercial rent is not the same as the more specific concept of economic rent, which refers to a payment made in excess of the payment needed to keep a factor of production, such as land, labour or capital, in current use. We shall return to this a little later.

The price of land

The special definitions of land and rent which are used by economists derive from those first used by David Ricardo in the 19th century. Ricardo was concerned, not with how much rent is paid for land used for a particular purpose, but how much rent is paid for land as a whole. He argued that:

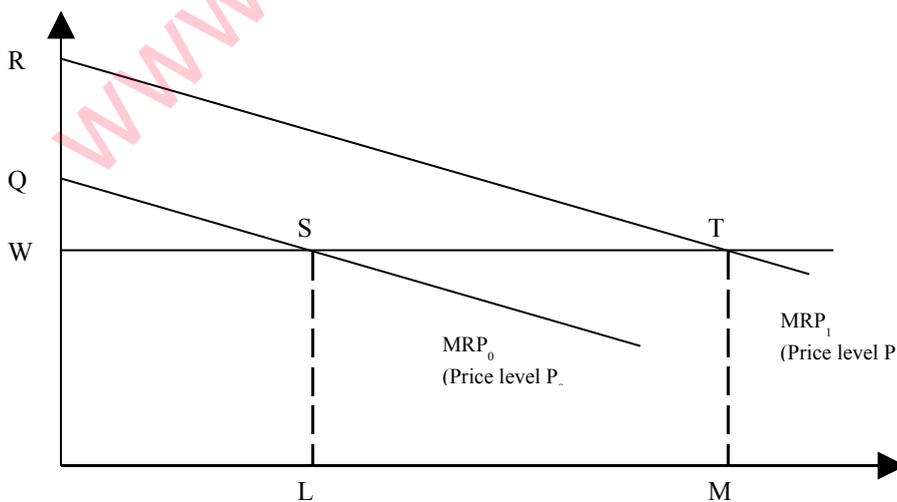
- (a) the total amount of land available is fixed, therefore the supply of land is inelastic, regardless of how much rent is paid for it;
- (b) since the supply of land is inelastic, the amount of rent will be determined by the price of the goods produced from the land for sale to markets.

Example: land and rent

As a simplified example, let us suppose that a piece of land (fixed in size) has only one use, which is to grow carrots. Suppose also that this fixed amount of land, which can only be used for growing carrots, has a perfectly elastic labour force to work on it and that no capital is employed.

Figure 11 shows:

- (a) the wage rate of labour, W , which is constant regardless of the number of workers employed, because labour supply is perfectly elastic;
- (b) the marginal revenue product of labour curves for two different price levels, P_0 and P_1 of a product, say carrots.



When the price of carrots is P_0 , OL workers be employed at the given wage rate:

- (i) the total cost of labour will be represented by the area $OWSL$;
- (ii) the total revenue from the sale of carrots will be $OQSL$;

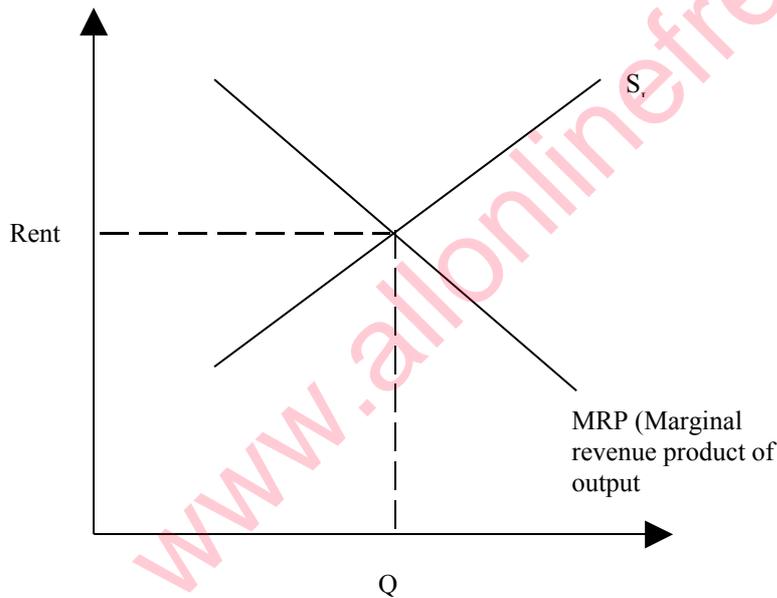
- (iii) the difference, WQS , is rent;
- (c) if the price of carrots rises to P , the MRP of labour will improve. OM workers will be employed, and;
 - (i) the total cost of labour is now $OWTM$;
 - (ii) the total revenue is $ORTM$
 - (iii) therefore rent has risen to WRT ;

We conclude that when land is in fixed supply, the size of the rent depends on the price of the goods produced on the land.

The price of land for specific uses

We can extend this principle to cover the situation where the supply of land is not perfectly elastic. This situation arises where we are considering the supply of land for a particular purpose, for example, office development or agricultural use. Since it is usually possible to change the use of a piece of land, the supply of land for a particular purpose should not normally be considered fixed.

If the price of agricultural land in the UK went up we would expect a transfer of land from (say) the domestic housing sector to the agricultural sector. In fact, the supply curve of agricultural land slopes upward from left to right. The equilibrium price and quantity of agricultural land are determined by the intersection of the demand curve (derived from the price of goods produced on the land) and the supply curve (derived from the price p which prevents a quantity of land q being transferred to any other use)



Entrepreneurship and profit

The function of the entrepreneur

The function of entrepreneurs, as the fourth factor of production, is to combine the other three factors, land, labour and capital, so as to maximise the efficiency of resource utilisation and to maximise the firm's profits. Business enterprise involves risk and uncertainty, and actual profits might be higher or lower than expected; and it is the entrepreneur's role to bear the burden of this uncertainty. His reward, profit, is what is left over after the other three factors of production have been rewarded, and if there is nothing left over the entrepreneur will make no profits or even a loss.

We can identify two aspects to the role of the entrepreneur;

- (a) he organises production and makes decisions about new business ventures;
- (b) he earns the reward of profit.

These dual aspects of entrepreneurship are possibly most apparent in partnerships and small private limited companies, where the owners of the business (partners or shareholders) are often also the senior managers. They organise production, make the decisions, and earn the profits for themselves.

The nature of profit

Unlike land, labour and capital, which are rewarded by rents, wages and interest respectively, the entrepreneur cannot be sure of gaining a reward (making a profit) because his business might make unanticipated losses.

Profit is the reward of the entrepreneur for the risks he takes. The entrepreneur bears the burden of business risk and uncertainty.

Normal profit

There ought to be an amount of 'pure' profit which an entrepreneur should expect as a reward for the risks that he or she takes. This expected or appropriate profit is known as 'normal' profit, and it is an economic cost of production.

Normal profit is earned when total revenues equal the total opportunity costs of all input resources. If revenues are just enough to equal opportunity costs, this means that the input resources are being used as well as they could be used anywhere else.

If actual economic profit is below normal profit (i.e. a 'loss' in economic terms) the firm would do better to leave the business it is in and put the resources at its disposal to better, more profitable use.

Supernormal profits

When total revenues exceed the total opportunity costs of input resources, the firm will be earning profit in excess of normal profits, and so resources are earning more than they could in an alternative occupation.

These excess profits are called supernormal profits, or monopoly profits (because monopoly firms are best able to exploit opportunities for earning supernormal profits). The supernormal profits of the monopolist are discussed in more detail in the next chapter.

Firms will obviously benefit from any supernormal profits that exist, and will wish to enjoy them if they are available. When a firm makes supernormal profits, however, other firms will want to enter the industry if they can, to grab a share of the high profits that are available. In competitive industries, supernormal profits therefore tend to be temporary, because they are eventually eroded by competition. Monopolies, however, might be able to earn supernormal profits indefinitely.

- (a) Entrepreneurs of better ability should always be able to earn supernormal profits when others are making only normal profits.
- (b) Monopolists can make supernormal profits by constructing entry barriers which prevent or deter rival firms from entering the market as competitors.
- (c) Supernormal profits can indicate to entrepreneurs the best markets for new investments or for a switch of their existing investments.

Normal profit and risk and uncertainty

Firms in high-risk industries should expect to earn a higher return than firms in low-risk industries. The higher the risk, the higher the 'risk premium' required. (Risk premium can be defined as the additional return in excess of a risk-free return needed to compensate an investor for making the risky investment).

When the opportunity costs of a business are measured, the opportunity cost of capital tied up (i.e. the opportunity cost of alternative investments) will be higher for higher risk firms.

Since 'normal profit' means a sufficient return to prevent the firm's owner from liquidating his investment and investing it elsewhere, normal profit must make allowance for the risk characteristics of the business, and what could be earned from an alternative investment of comparable risk. Thus, normal profit will be higher in a more risky market than in a less risky market.

www.allonlinefree.com

Market Structures: From Perfect Competition To Monopoly

Your objectives

After completing this chapter you should:

- (a) understand the range of market structures that may exist;
- (b) understand the main features of two extreme market structures: perfect competition and monopoly.

Perfect and imperfect competition

In Chapter 2 we set out the characteristics of a hypothetical perfect market. Consideration of such a market is useful despite the fact that it never exists in practice. The reason for this is that the underlying theory can be studied without the complicating factors that make real-life (imperfect) markets so complex.

In this chapter we look at both perfect competition and its opposite, monopoly, and examine the effects of these market structures on the output decisions of firms. In the next chapter we look at two less extreme market structures: monopolistic competition and oligopoly.

Characteristics of imperfect competition

One or more of a number of factors may make competition conditions 'imperfect'

- (a) There may be one or just a few large firms dominating the market.
 - (i) A single firm dominating the market is called a monopolist, or monopoly firm.
 - (ii) When two or several firms jointly dominate the market, there is an oligopoly and firms are referred to as oligopolists. (A special case of oligopoly is duopoly: this is when precisely two firms dominate the market.)

When only one or a few firms dominate the market, most of the market demand will be for their products and the firm can have a significant influence on the market price. The firm can make decisions on price as well as output and the firm is a price maker .

- (b) There may be shortages of information about prices and profit opportunities among some consumers or producers.
- (c) In some markets, firms succeed in creating differences, real or imagined, between their own product and similar products of competitors. They seek to create customer demand for their own products in preference to competitors' products by emphasising these differences. Two ways of creating differences are branding and advertising . Another is small variations in product design. Creating such differences is called product differentiation. and it enables firms to become like monopolists or oligopolists in their own special corner of a large market. This type of competition is referred to as monopolistic competition.
- (d) Free entry into the market might be prevented by existing firms in the market.

Equilibrium under perfect competition

Equilibrium in the short run

How are price, output and the maximisation of profit determined in the case of the firm operating under conditions of perfect competition in the short run?

The short run refers to a period in which the number of firms in the market is temporarily fixed. In these circumstances it is possible for firms to make supernormal profits or losses as the following diagrams show.

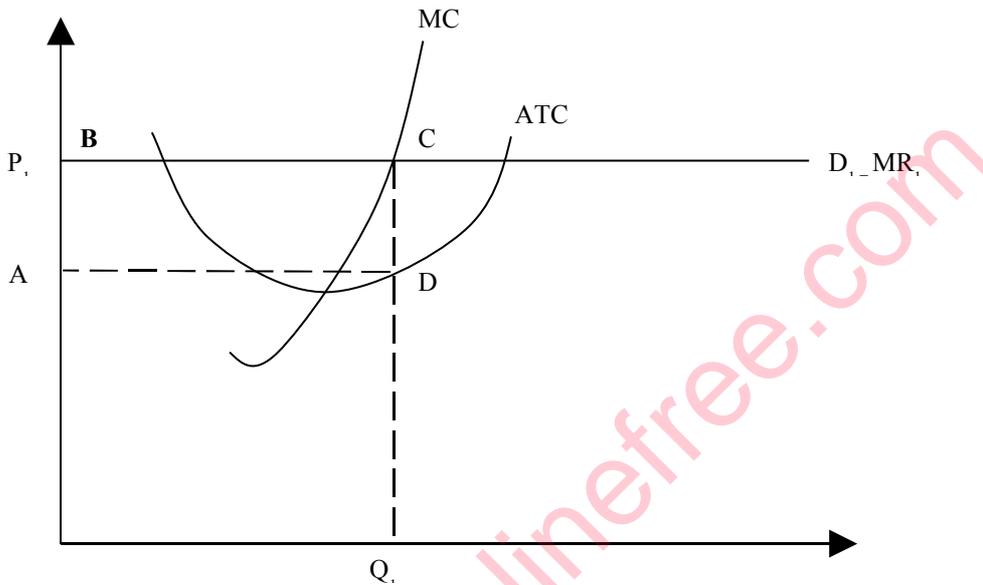
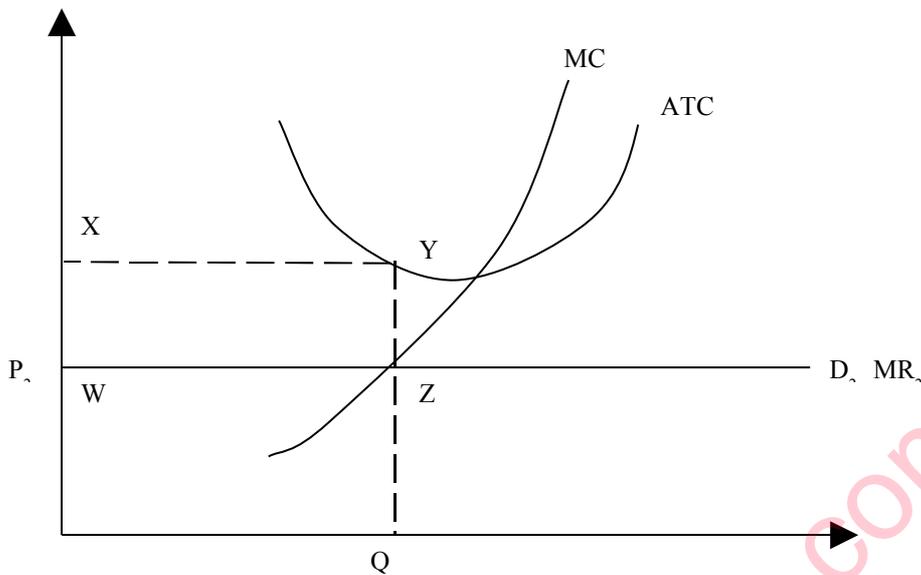


Figure 1 shows the cost and demand curves of a firm in the short run making supernormal profits. The demand curve is the horizontal line D_1 at price P_1 . The curve is a horizontal line indicating that the firm may not influence the price of the goods and has to accept the price that the market as a whole fixes for them. If the firm were to charge a higher price it would lose all its sales and there is no point charging a lower price as it can sell all its output at the given price. The demand curve is thus also the marginal revenue curve; every new unit sold at price P_1 increases total revenue by an amount P_1 .

Figure 1 also shows the average total cost curve (ATC) and the marginal cost curve (MC), with the MC cutting the ATC at the lowest point of the ATC. Given these cost curves and the demand curve D_1 , the firm will produce the output Q_1 , where the MC curves cuts the MR horizontal curve at the point C. This is the profit maximising point.

If the firm were to produce fewer units it would be producing at a point where MR was higher than MC and all additional units produced up to the point where $MR = MC$ would similarly have MR greater than MC. The firm should produce these units because so long as MR is greater than MC, each unit shows a profit (additional revenue is greater than additional costs). Similarly it should not expand production past $MR = MC$ because it will be producing where MC is greater than MR, in other words where the additional costs for each unit exceed the additional revenue earned.

At the output Q_1 the firm is making supernormal profits indicated by the rectangle ABCD. This will attract new firms into the industry and the price will be bid down, possibly to price P_2 as shown in Figure 2. Here the firm makes a loss shown by the rectangle WXYZ. Once again the firm produces where $MC = MR$ giving an output of Q_2 .

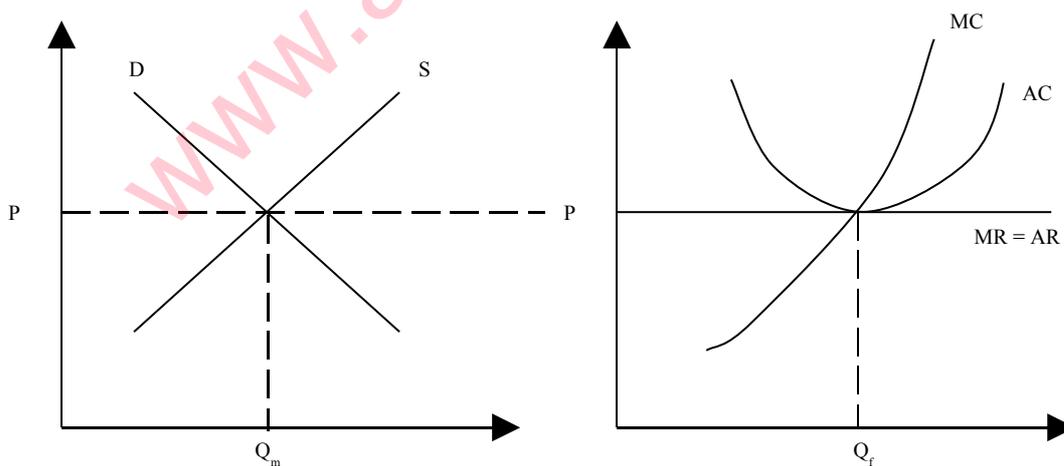


In the long run, whenever profits are being made new firms will enter the industry and the price will fall. Similarly, when losses are made firms will leave the industry and the price will rise.

Equilibrium in the long run

In a perfectly competitive market in the long run, the firm cannot influence the market price and its average revenue curve is horizontal. The firm's average cost curve is u-shaped. The firm is in equilibrium and earns normal profits only (i.e. no supernormal profits) when the AC curve is at a tangent to the AR curve as shown in Figure 1(b). In other words, long-term equilibrium will exist when supernormal profits and losses are eliminated.

There is no incentive for firms to enter or leave the industry and the price will remain at p with the firm making normal profits only.



Note the following points about Figure 3.

- (a) The market price p is the price which all individual firms in the market must take,
- (b) If the firm must accept a given MR (as it must in conditions of perfect competition) and it sets $MR = MC$, then the MC curve is in effect the individual firm's supply curve (Figure 3(b)). The

market supply curve in Figure 3(a) is derived by aggregating the individual supply curves of every firm in the industry.

Long-run equilibrium will, then, occur in the industry when there are no more firms entering or leaving the industry because no new firm thinks it could earn higher profits by entering and no existing firm thinks it could do better by leaving. In the long run, then, all firms in the industry will have $MR = MC = AR$ as in Figure 3(b).

Equilibrium for a monopoly

The monopoly market

The long-run equilibrium of a firm in a perfectly competitive market is at a price and output level where only normal profits are earned. $Price = MR = MC = AC$, and the firm produces at the minimum average cost per unit.

Monopoly is the converse to perfect competition. In a monopoly, there is only one firm, the sole producer of a good which has no closely competing substitutes, so that the total market supply is identical with the single firm's supply.

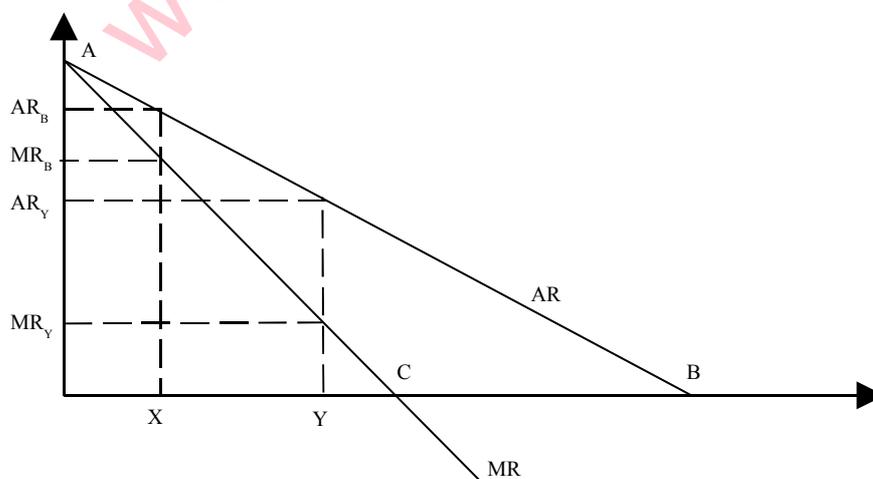
In monopoly the firm faces a downward-sloping average revenue curve because its average revenue curve is the same as the total market demand curve.

If average revenue is falling, marginal revenue will always be lower than average revenue; if the monopolist increases output by one unit the price per unit received will fall, so the extra revenue generated by the sale of the extra unit of the good is less than the price of that unit. The monopolist therefore faces a downward-sloping AR curve with an MR curve below the AR curve.

There is a useful technique for drawing the AR 'curve' and the MR 'curve' of a monopolist.

- For simplicity, assume that the monopolist's AR curve, which is the market demand curve too, is a straight line.
- Draw this AR curve from a point on the y axis (point A in Figure 4 below) to a point on the x axis (point B in Figure 4).

You can now draw the MR curve from the same point on the y axis (point A) to a point on the x axis which is half way to point B. This is shown as point C in Figure 4, and $OC = CB$ (i.e. $OC = \frac{1}{2}OB$).



The marginal revenue can be negative. This is when the price elasticity of demand is inelastic and so although lowering the price increases sales demand, the volume increase is small and so total revenue falls.

It is obviously important that you should understand what the MR and AR (demand) curves are showing us. In Figure 4:

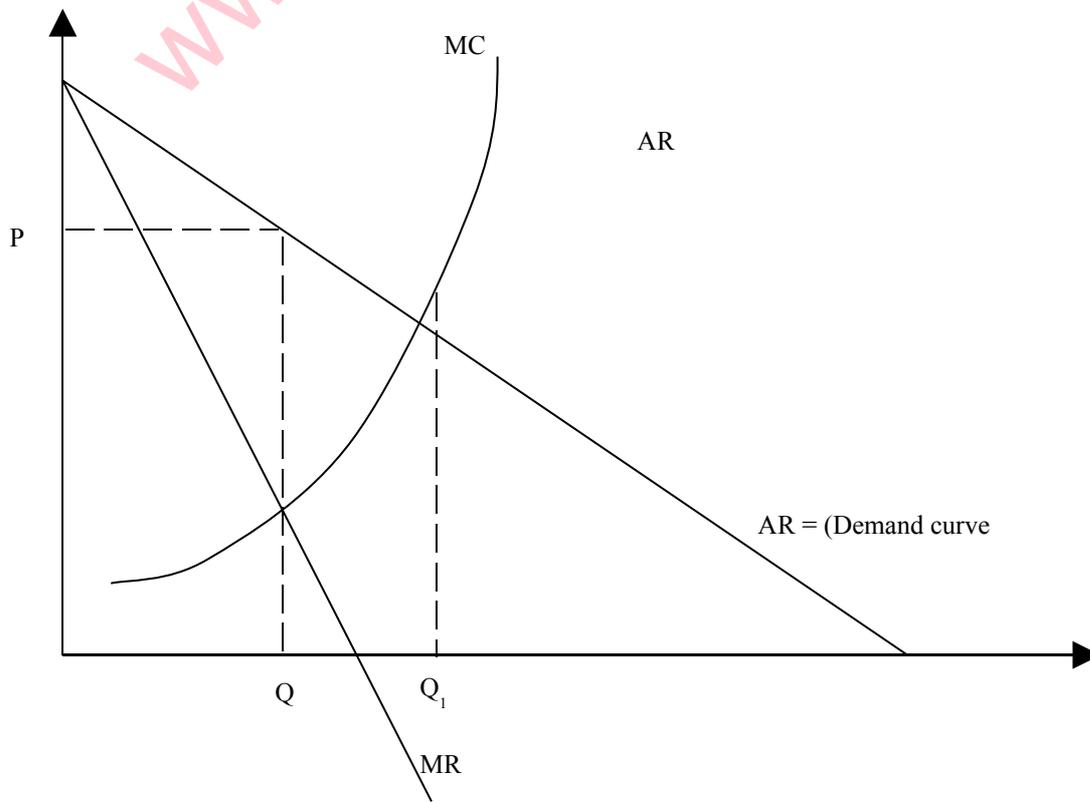
- (a) at output quantity X , the marginal revenue earned from the last unit produced and sold is MR_X , but the price at which all the X units would be sold is P_X . This is found by looking at the price level on the AR curve associated with output X ;
- (b) similarly, at output quantity Y , the marginal revenue from the last unit produced and sold is MR_Y , but the price at which all y units would be sold on the market is, from the AR curve for y output, P_y .

Profit-maximising equilibrium of a monopoly

The condition for profit maximisation is, as we have seen, that marginal revenue should equal marginal cost. This is true for any firm. As long as marginal revenue exceeds marginal cost, an increase in output will add more to revenues than to costs, and therefore increase profits. A monopolist might maximise profits

- (a) but make no supernormal profits; or
- (b) make supernormal profits.

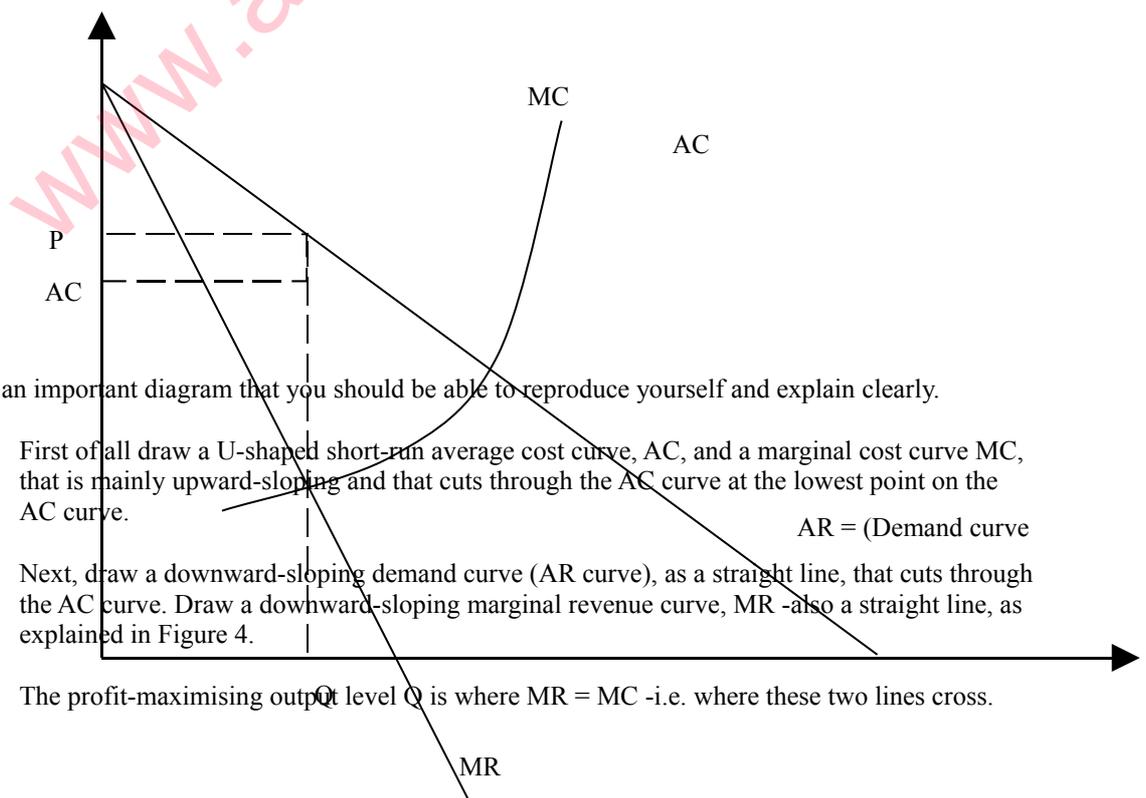
Figure 5 shows a monopoly equilibrium where the monopolist is earning just normal profits, and so $AC = AR$. At this point (Q) the AC curve touches the AR curve at a tangent, at exactly the same output level where $MC = MR$. Since $AC = AR$ and AC includes normal profits, the monopolist will be earning normal profits but no supernormal profits in this situation.



In this situation, the monopoly will make a loss by producing at output higher than Q_1 and so it will have to produce at an output level which is below the capacity at which its average costs are minimised (output Q_1).

Monopolies are usually able to earn 'monopoly' or supernormal profits in the long term as well as the short term, and the situation illustrated in Figure 5 will be rare for a monopoly, although (as we shall see later) it is a long run equilibrium situation for firms in 'monopolistic competition'. It might, however, represent the long-term equilibrium of some monopolies, where barriers to entry into the industry are low, and competition would be attracted into the market if supernormal profits were achievable. It might also be the case for a government-run monopoly which is required to cover costs rather than being required to make profits.

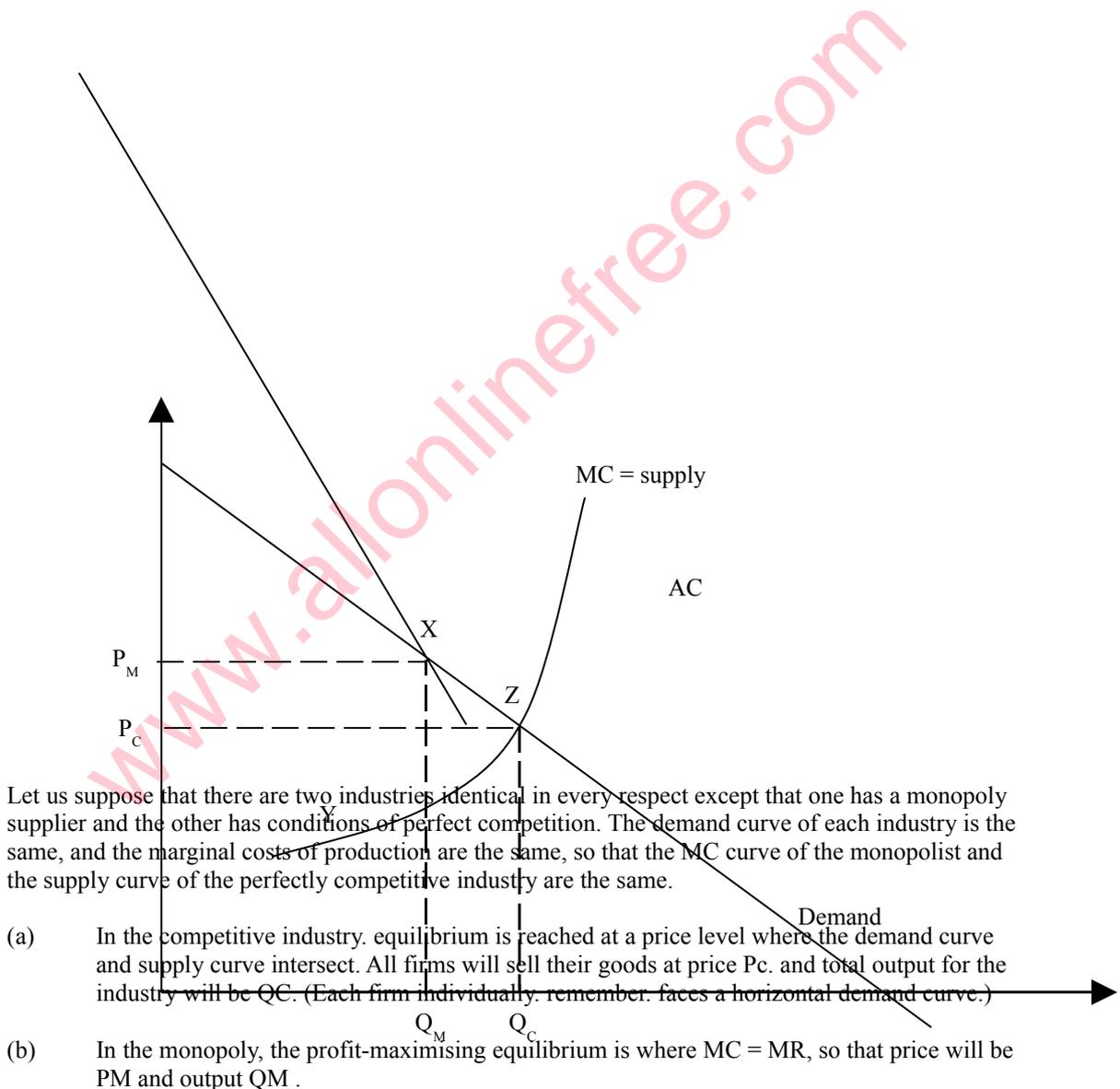
In perfect competition, a firm should not be able to earn supernormal profits in the long run because they would be 'competed away' by new entrants to the industry. In monopoly, however, the firm can earn super-normal profits in the long run as well as in the short run, because there are barriers to entry which prevent rivals entering the market. Figure 6 shows a firm earning supernormal profits, equal to the shaded area, which represents $(P - AC) \times Q$ units produced and sold



- (d) The profit-maximising selling price p is found by reference to the demand curve (i.e. AR curve) at that output level.
- (e) The average cost at this output level, AC , is found by reference to the AC curve.
- (f) You will find that at this level of output, the price, AR , exceeds average costs, AC , and so super-normal profits are earned (= unit price minus average cost \times volume of output in units).

Monopoly and perfect competition compared

If there are no economies of scale for a large output-volume firm, then output will be lower, and costs and price higher, under monopoly as compared with perfect competition.



Thus, a firm in a monopoly industry will produce less and sell at a higher price than firms in a perfectly competitive industry.

It is also significant that:

- (a) in the perfectly competitive industry, equilibrium is reached when firms minimise their average costs (AC is at a minimum). Position Z is then the output/price level that gives an optimal allocation of resources;
- (b) in the monopoly industry, profit-maximising equilibrium is not at the output level where AC is minimised (but at a lower output level with a higher AC).

At the monopolist's profit-maximising output, the difference between price and marginal cost is XY. It could be argued that if the firm were to produce more output, it would provide more marginal utility for consumers (as represented by the demand curve) than it would create marginal cost for the firm. There is a social cost that occurs because a monopoly produces only QM of output at price \sim , rather than QC of output at price P_c . The loss of social welfare, or social cost, is represented by the area XYZ in Figure 7, and is sometimes referred to as the deadweight burden of monopoly.

These conclusions might not be valid, however, if the monopolist is able to take advantage of its monopoly position to achieve economies of scale that perfectly competitive firms could not achieve, owing to their relative smallness in terms of the total market output.

This is an important proviso, because it is a theoretical argument in favour of monopolies.

www.allonlinefree.com

Price discrimination

Price discrimination and market segmentation

The term price discrimination refers to a situation in which a firm sells the 'same' product at different prices in different markets.

Market segmentation may involve elements of product differentiation (e.g. different brand names) in order to satisfy particular segments of the market.

Three basic conditions are necessary for price discrimination to be effective and profitable.

- The seller must be able to control the supply of the product. Clearly, this will apply under monopoly conditions. The monopoly seller has control over the quantity of the product offered to a particular buyer .
- The seller must be able to prevent the resale of the good by one buyer to another. The markets must, therefore, be clearly separated so that those paying lower prices cannot resell to those paying higher prices. The ability to prevent resale tends to be associated with the character of the product, or the ability to classify buyers into readily identifiable groups. Services are less easily resold than goods while transportation costs, tariff barriers or import quotas may separate classes of buyers geographically and thus make price discrimination possible.
- There must be significant differences in the willingness to pay among the different classes of buyers. In effect this means that the elasticity of demand must be different in at least two of the separate markets so that total profits may be increased by charging different prices

We can see how the monopolist seller practising price discrimination can maximise revenue using a diagrammatic illustration (Figure 8).

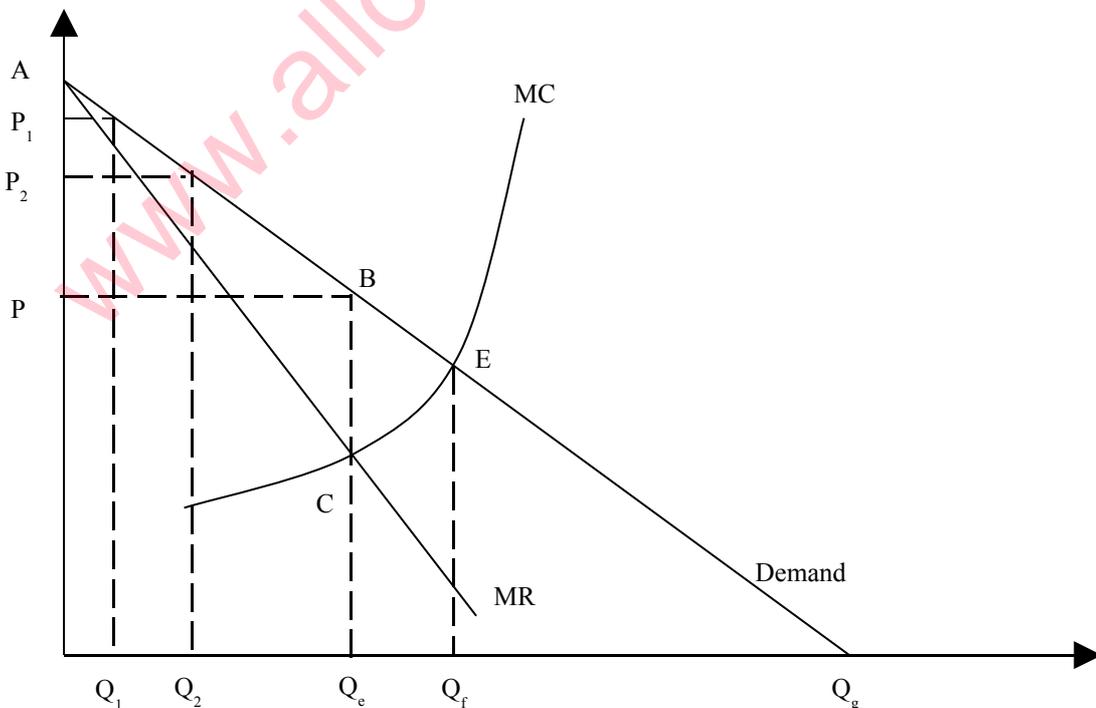


Figure 8 demonstrates firstly the equilibrium position of a monopolist that does not discriminate. He produces at the point C where marginal cost equals marginal revenue, producing output Q_e and selling at price P. His total revenue is given by the rectangle $OPBQ_e$. Figure 8 also illustrates how the

discriminating monopolist can improve on this position, both from the point of increased revenue and increased profits.

The discriminating monopolist does not charge the same price for all units sold. If we assume that the monopolist can discriminate perfectly, then he can sell each unit for a different price as indicated on the demand curve. Thus he can sell the first unit Q_1 at the price P_1 , and the second unit Q_2 at the price P_2 . This follows for all units sold so that the demand curve becomes the marginal revenue curve; each extra unit sold is sold for the price indicated on the demand curve, each previous unit being sold for the higher price relevant to that unit.

The perfectly discriminating monopolist will still maximise profits where $MC = MR$, but the marginal revenue curve is now the curve D , the demand curve. He thus produces at the point E where marginal cost equals the new marginal revenue, producing Q_f units.

The additional revenue of the discriminating monopolist is represented by the areas APB plus Q_eBEQ_f . The discriminating monopolist has thus maximised his revenue (consistent of course with maximising his profit). If the monopolist did not wish to maximise profit but wished simply to maximise revenue he would expand production to the point Q_g when his total revenue would be the area OAQ_g .

Take care not to confuse maximising revenue with maximising profit. Increasing output beyond Q_f in the example will not increase profit as marginal costs exceed marginal revenue for each additional unit sold.

Price discrimination often results where the market is separated by transport costs and tariffs; hence firms may sell their products abroad cheaper than at home. An extreme example of this is known as dumping; this occurs when exports sell in foreign markets at prices below the cost of production. This pricing strategy is designed to drive domestic producers out of the industry so that the foreign producer may achieve monopoly power. The price differential cannot, of course, exceed the cost of transporting the good back to the home market plus any tariff on imports.

Price discrimination also occurs where it is possible to separate buyers into clearly defined groups. Industrial users of gas and electricity are able to purchase these fuels more cheaply from British Gas and electricity companies than are domestic users. Similarly milk is sold more cheaply to industrial users, for example for making into cheese or ice cream, than to private households.

Public policy towards monopolies

Public policy towards private enterprise monopolies

Monopolies might be harmful or beneficial to the public interest.

- (a) A beneficial monopoly is one that succeeds in achieving economies of scale in an industry where the minimum efficiency scale is at a level of production that would mean having to achieve a large share of the total market supply.
- (b) A monopoly against the public interest would be one in which cost efficiencies are not achieved, or are negligible. Oliver Williamson suggested that inefficiency in monopolies might occur if 'market power provides the firm with the opportunity to pursue a variety of other-than-profit objectives.'

There are other reasons for trying to control monopoly growth. A monopoly firm may be a multinational, with its head office in another country. Multinational firms are difficult to control within the context of a government's economic policy requirements and so a government might prefer to see more 'national' firms in a position of some strength in the country's home markets.

Monopolies might also try to preserve their monopoly position by acting to prevent competition. They might create barriers to entry into the industry against other potential rivals (e.g. by taking over small

new competitors or purchasing patents to secure a production monopoly). When competition between firms is 'killed off' in this way, the public interest is harmed.

A monopoly may try to exploit its position by using price discrimination, to charge higher prices for the same good in a different segment of the market. Certain pharmaceutical manufacturers, for example, have been accused of unfairly charging higher prices in one country than in another. This too can be against the public interest.

Government control over monopolies, mergers and restrictive practices

There are several different ways in which a government can attempt to control monopolies.

- (a) It can stop them from developing, or it can break them up once they have been created. In the past, there has been a history of 'trust-busting' in the USA. Preventing monopolies from being created is the reason why a government might have a public policy on mergers.
- (b) It can take them over. Nationalised industries are often government-run monopolies, and central and/or local government also have virtual monopolies in the supply of other services, such as health, the police, education and social services. Government-run monopolies are potentially advantageous because:
 - (i) they need not have a profit-maximising objective so that the government can decide whether or not to supply a good or service to a household on grounds other than cost or profit;
 - (ii) the government can regulate the quality of the good or service provided more easily than if the industry were operated by private firms;
 - (iii) key industries can be protected (e.g. health, education).
- (c) It can allow monopolies or oligopolies to operate, but try to control their activities in order to protect the consumer. For example, it can try to prohibit the worst forms of restrictive practice, such as price cartels. Or it may set up regulatory bodies to protect consumers' interests where conditions of natural monopoly apply, as in the recently privatised utility industries of the UK. The pricing of products in such industries may be controlled: in the UK, many of the large utilities have been required to limit price increases to a specified percentage below the Retail Prices Index over a number of years.

There are two basic tenets in the thinking behind consumer protection policies.

- (a) Control over markets can arise by firms eliminating the opposition, either by merging with or taking over rivals or stopping other firms from entering the market. The problem here is that when a single firm controls a big enough share of the market it can begin to behave as a monopolist even though its market share is below 100%.
- (b) Several firms could behave as monopolists by agreeing with each other not to compete. This could be done in a variety of ways (for example by exchanging information, by setting common prices or by splitting up the market into geographical areas and operating only within allocated boundaries).

In a perfect monopoly, there is only one firm that is the sole producer of a good that has no closely competing substitutes, so that the firm controls the supply of the good to the market. The definition of a monopoly in practice is rather more extensive than this, because governments seeking to control the growth of monopoly firms will probably choose to regard any firm that acquires a certain share of the market as a potential monopolist.

The Monopolies and Mergers Commission in the UK

In the UK, the Monopolies and Restrictive Practices Act 1948 provided that any firm controlling more than one third of the market for its goods should be investigated as a potential monopoly which was against the public interest. Under the Fair Trading Act 1973, the Director General of Fair Trading is

allowed to refer cases to the Monopolies and Mergers Commission if any firm controls one quarter of the market, or if any proposed takeover or merger would create a firm that controlled more than one quarter of the market. The Commission will then investigate the proposed merger or takeover and recommend whether or not it should be allowed to proceed.

Another aspect of the work of the Monopolies and Mergers Commission (MMC) is to investigate cases where a monopoly is suspected of operating against the public interest and to recommend to the government the steps that should be taken to make the monopoly alter its practices. In 1991, an inquiry into the carbonated drinks industry was published. At the time of writing (August 1993) an inquiry is expected into the music industry, prompted by a widespread feeling that compact discs are priced too highly in the UK.

UK government policy on monopolies and mergers

It is significant, however, that a government department, the Office of the Director General of Fair Trading has the power to refer cases to the Commission, which can then prevent a merger or takeover from taking place if it is considered to be against the public interest. The strength of anti-monopoly and anti-takeover activity therefore depends on the attitude of the government of the day, as expressed by the activities of the office of the Director General of Fair Trading.

The office of the Director General has the help of the Consumer Protection Advisory Committee which advises the Director General and the Minister whether trading and commercial practices referred to it are harmful to the interests of the consumers. (The activities of this Committee are not directed against monopolies as such, but it is a focus of consumer protection activity which is one of the reasons why anti-monopoly controls might be required.)

Mergers

Mergers, or prospective mergers, can also be referred to the MMC if they are thought likely to create or strengthen a monopoly. Companies do not have to seek permission before merging, nor can permission be given in advance of the full details being known. But it is common practice these days for mergers to be proposed subject to MMC approval.

Public policy in favour of monopolies and mergers

One aspect of the UK government's policy has been an attempt to control the development of monopolies that are 'against the public interest'. But opinions differ about what monopolies would be good for the country and which ones would be harmful. Another aspect in recent years has been the establishment of 'consumer watchdog' bodies to regulate the recently privatised utility industries where there are usually conditions of natural monopoly.

In the modern world of multinational companies, companies need to be big to survive and prosper. Arguably, the UK's industrial strength has declined over the years because the country has failed to nurture enough multinational companies (strengthened by virtual monopolies in their own country) to compete successfully in world markets. Compare Rover cars, as just one example, a very small car manufacturer in world terms, with Fiat of Italy, the largest company in Italy, a big multinational with a virtual monopoly in the Italian car market.

There have been signs in recent years that the UK government has wanted to encourage the growth of companies in the UK, and many of the proposed 'mega-mergers' (i.e. mergers between big companies) have not been referred to the Monopolies and Mergers Commission. Government anti-monopoly policy in the UK has accordingly been relatively mild in recent years.

European Community regulations on mergers

Under a regulation introduced during 1990, the European Commission has the power to intervene and to either block or to authorise larger mergers.

The regulation provides that mergers with a world wide aggregate turnover of over 5 billion ecu (approximately £3.5 billion) between undertakings with EC-wide turnover of over 250 million ecu (approximately £175 million) require prior clearance from the European Commission. If the Commission finds that the merger raises serious doubts as to its compatibility with the European common market, it will initiate proceedings to block the merger.

Privatisation of nationalised industries

Privatisation takes three broad forms.

- (a) The deregulation of industries, to allow private firms to compete against state-owned businesses where they were not allowed to compete before (e.g. deregulation of bus and coach services, possible deregulation of postal services).
- (b) Contracting out work to private firms, where the work was previously done by government employees (e.g. refuse collection, hospital laundry work).
- (c) Transferring the ownership of assets from the state to private shareholders (e.g. the denationalisation of British Telecom and British Gas).

The UK government has carried out a policy of denationalisation in recent years. British Gas, British Telecom and the main Water Authorities have been among the enterprises which have been privatised.

Privatisation may improve efficiency, but there are other possible advantages of privatisation.

- (a) Denationalisation provides an immediate source of money for the government.
- (b) Privatisation reduces bureaucratic and political meddling in the industries concerned.
- (c) There is a view that wider share ownership should be encouraged. Denationalisation is one way of creating wider share ownership, as the sale of shares in British Telecom, British Gas and some other nationalised industries proved in the UK.

There are arguments against privatisation too.

- (a) State-owned industries are more likely to respond to the public interest, ahead of the profit motive. For example, state-owned industries are more likely to cross-subsidise unprofitable operations from profitable ones; e.g. the Post Office will continue to deliver letters to the isles of Scotland even though the service might be very unprofitable.
- (b) Encouraging private competition to state-run industries might be inadvisable where significant economies of scale can be achieved by monopoly operations.

The advantages of having a controlled or centrally planned economy, rather than a free enterprise economy, can also be advanced as reasons in favour of having nationalised industries.

Monopolistic Competition And Oligopoly

Your Objectives

After completing this chapter you should:

- (a) understand the main features of monopolistic competition and their implications for firms marketing and output decisions.
- (b) understand the main features of oligopoly and their implications for firms marketing and output decisions.

Monopolistic competition and non-price competition

Monopolistic competition

Actual market structures usually do not correspond with either of the extreme cases of perfect competition and monopoly. This leads to a discussion of forms of imperfect market structure, including a form known as 'monopolistic competition'.

Firms in monopolistic competition (as well as oligopoly, discussed later in this chapter) will try to avoid competition on price in order to preserve their position as a price maker. They will often resort to non-price competition instead. This can take several forms, including:

- (a) product differentiation;
- (b) branding;
- (c) advertising and sales promotion;
- (d) creating 'add-on' services

Product differentiation

Product differentiation describes a situation in which there is a single product being manufactured by several suppliers, and the product of each supplier is basically the same. However, the suppliers try to create differences between their own product and the products of their rivals or between different brands of the product which they sell. These differences might be real (e.g. small or large design differences) or imaginary, created largely by advertising and brand image.

Differentiation may take a number of forms, including:

- (a) different physical or technical characteristics, satisfying different buyer needs, or the same needs in different ways;
- (b) different packaging;
- (c) different conditions of sale with respect to guarantees, after-sales services, and so on;
- (d) different geographical location;
- (e) different perceptions of the product created through advertising and promotion. This includes branding of goods.

Other forms of non-price competition

There are other ways in which non-price competition can be developed. As the term implies, non-price competition involves competition between firms for customers in the same market, but not on the basis of lowest price (or, at least, not on the basis of low price only).

- (a) Advertising and sales promotion, or brand imagery
 - (i) The aim of advertising, sales promotions or brand image is to increase demand for the good, often at the expense of demand for the goods of other firms in the market.
 - (ii) When one firm advertises, others will retaliate. As a result, an advertising or sales promotion campaign might have little or no effect on demand, and so incur cost for little or no benefit.
- (b) Incidental services. Incidental services are 'extra' services that come as an 'add-on' to the basic good. A firm might make its product more attractive than its rivals' by means of, for example, superior sales services or personal attention. Rival firms can retaliate by offering incidental services too.
- (c) Innovation and technical differences. A firm may enhance its market share by creating genuine differences through innovation: Brand X is made to have more 'features' than Brand Y, so that customers for Brand Y are encouraged to switch. Here, the opportunity for retaliation by other firms must be limited because innovation is easily the most effective form of non-price competition; however, it is not easily available to firms which produce a basic commodity, such as petrol or farm produce. e.g. milk or wheat.

Profit-maximising equilibrium of a firm in monopolistic competition

A firm which operates in conditions of monopolistic competition:

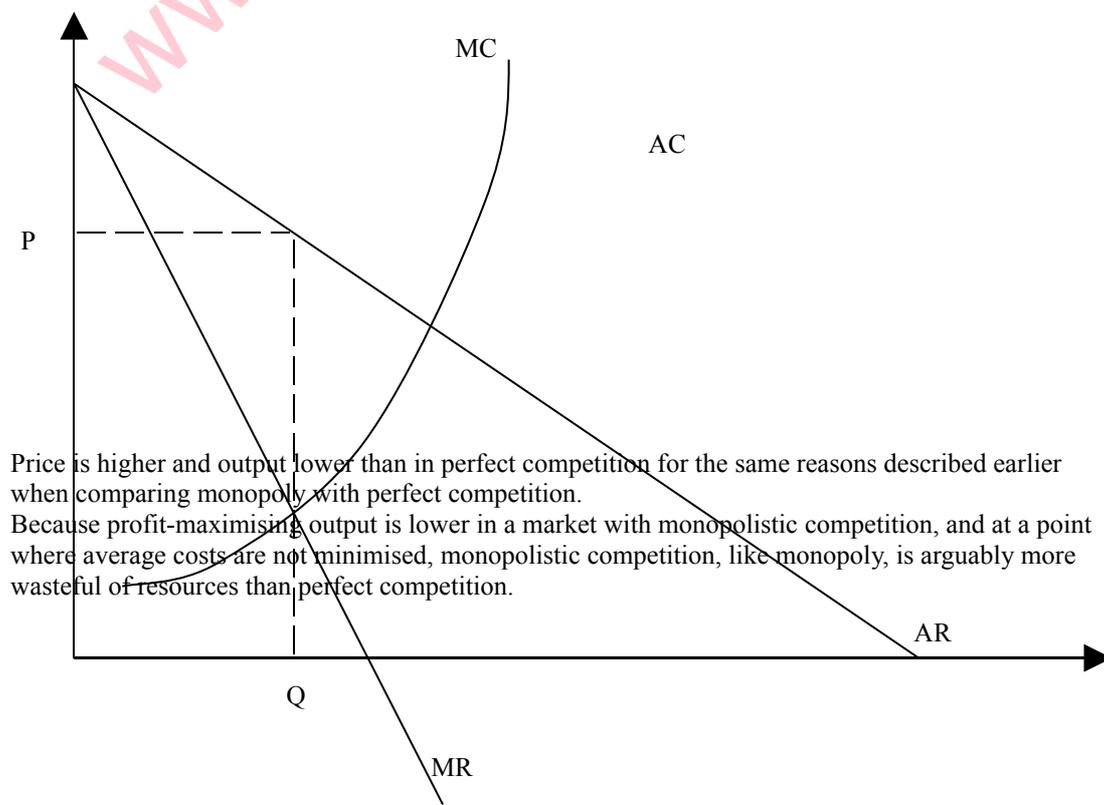
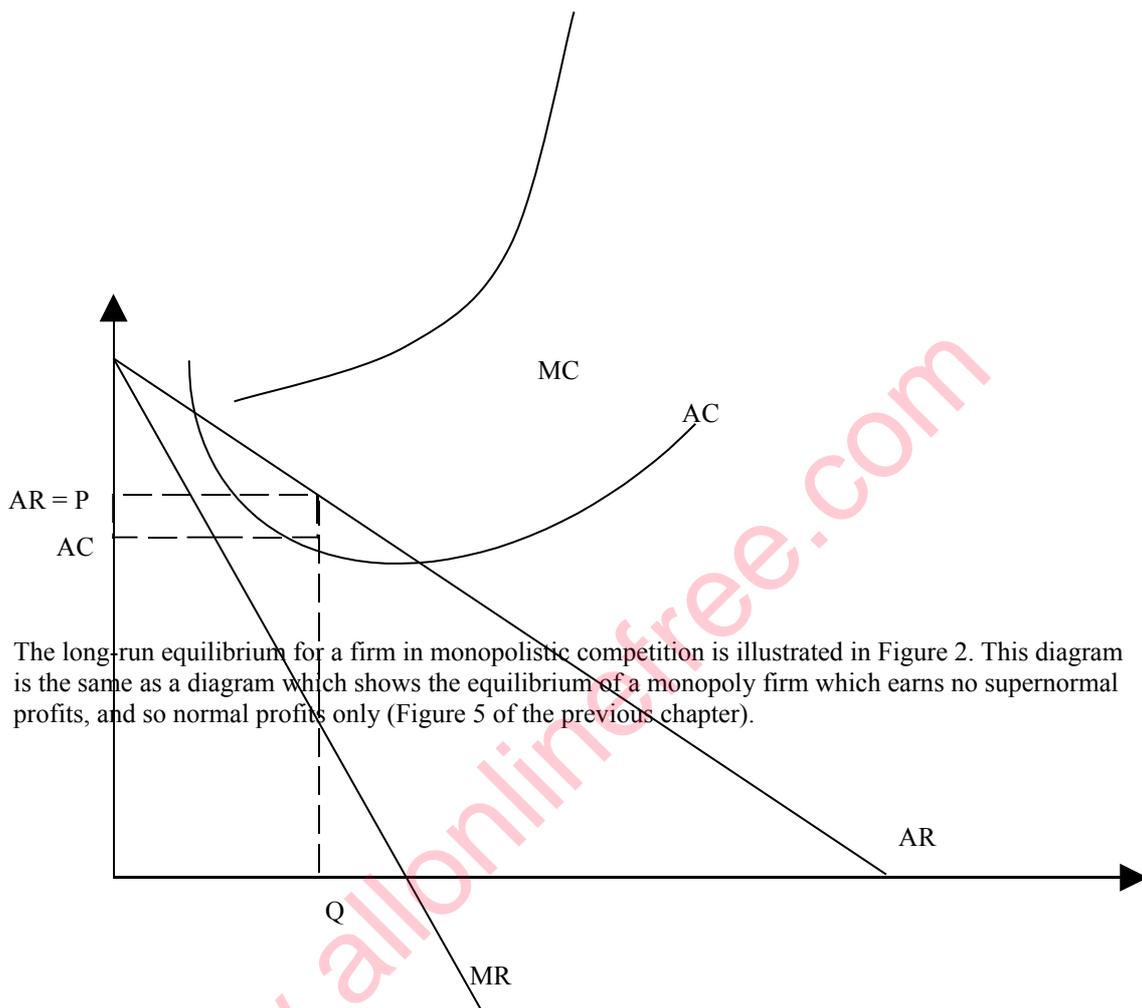
- (a) has a downward-sloping demand curve. The quantity demanded responds to the price at which the firm is prepared to sell. The downward-sloping curve is possible because of product differentiation;
- (b) unlike a monopoly firm, is often unable to create barriers to entry to other firms. (Indeed, the firm already competes with rivals, which can take retaliatory competitive action if the firm makes big profits).

A firm will therefore have:

- (a) a short-run equilibrium, in which it can make supernormal profits; and
- (b) a long-run equilibrium.

In the long-run, the monopolistic competitor cannot earn supernormal profits since there are no entry barriers. Its short-run supernormal profits will be competed away by new entrants. So the firm will eventually be able to achieve normal profits only.

The short-run equilibrium for a firm in monopolistic competition is illustrated in Figure I below. This diagram is the same as a diagram which shows the equilibrium of a monopoly firm earning supernormal profits. The firm makes supernormal profits of $(P-AC) \times Q$ units, shown by the shaded area.



Characteristics of monopolistic competition: a summary

Monopolistic competition is characterised by the following.

- (a) Equilibrium output below the level at which average costs are a minimum, in the long-run. Monopolistic competitors are said to have excess or unused capacity.
- (b) Since firms in monopolistic competition cannot expand their output to the level of minimum average cost output without making a loss, the excess capacity theorem predicts that industries marked by monopolistic competition will always tend to have excess capacity. Check this in Figure 2, where profit is maximised at output Q , and output Q is lower than the output level where AC would be minimised.
- (c) A higher price and lower output than in perfect competition.
- (d) An equilibrium price above marginal cost.
- (e) Product differentiation.
- (f) Other forms of non-price competition such as advertising and sales promotion.

Is monopolistic competition wasteful of economic resources?

There are several reasons for suggesting that monopolistic competition is wasteful of economic resources.

- (a) It can be argued that it is wasteful to produce a wide variety of differentiated versions of the same product. If a single version of the same product were made, firms might be able to achieve economies of scale with large-volume production (and so shift their cost curves to the right).
- (b) Some methods that are used to create product differentiation are a waste of resources. Advertising costs are arguably an example of this, although some would argue that promotional activity is informative to potential customers and actually adds utility to a product.
- (c) Firms in monopolistic competition, like monopolists, produce at an output level below that at which where AC is minimised.

Oligopoly

The nature of oligopoly

Oligopoly differs from monopoly in that there is more than one firm in a market, although the number of firms is small. Oligopoly differs from monopolistic competition because in oligopoly the number of rival firms is small.

An oligopoly consisting of only two firms is sometimes referred to as a duopoly.

Oligopolists may produce a homogeneous product (eg oil) or there may be product differentiation (eg cigarettes, cars).

The essence of oligopoly is that firms' production decisions are inter-dependent. One firm cannot set price and output without considering how its rivals' response will affect its own profits. How an oligopolist will actually set his output and price depends on what assumption firms make about their competitors' behaviour.

Price cartels by oligopolist producers

A price cartel or price ring is created when a group of oligopoly firms combine to agree on a price at which they will sell their product to the market. The market might be willing to demand more of the product at a lower price, while the cartel agreement attempts to impose a higher price (for higher unit profits) by restricting supply to the market to a level which is consistent with the volume of demand at the price they wish to charge.

Each oligopoly firm could increase its profits if all the big firms in the market charge the same price as a monopolist would, and split the output between them. This is known as collusion, which can be tacit or openly admitted.

Cartels are difficult to outlaw. There might still be collusive price leadership. This occurs when all firms realise that one of them is initiating a price change that will be of benefit to them all, and so follow the leader and change their own price in the same way.

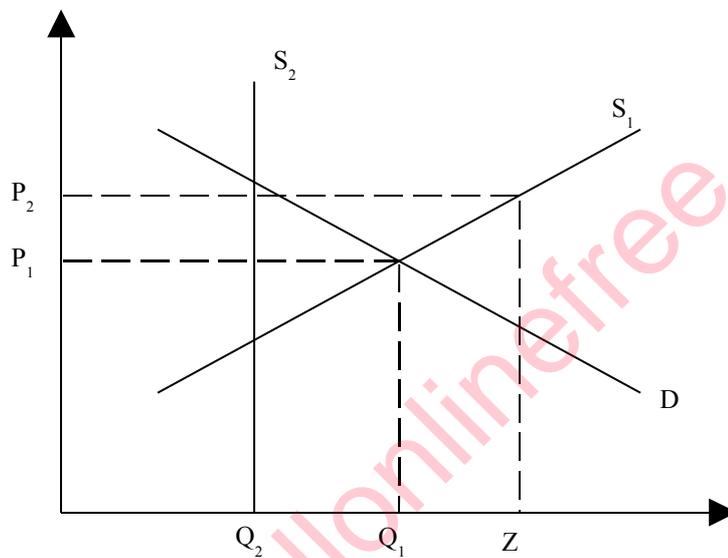


Figure 3 shows that:

- in a competitive market, with a market supply curve S_1 and demand curve D , the price would be P_1 and output Q_1 ;
- a cartel of producers might agree to fix the market price at p_2 , higher than p_1 . But to do so, the cartel must also agree to cut market supply from Q_1 to Q_2 , and so fix the market supply curve at S_2 .

Establishing a cartel depends on:

- the firms in the cartel being able to control supply to the market;
- agreeing on a price (P_2);
- agreeing on how much of the output each firm should produce. In Figure 3, if the market price is fixed at P_2 , firms would want to supply output Z in a free market. This cannot be allowed to happen; otherwise market price P_2 couldn't be sustained.

The main weakness with cartels is that each firm is still seeking the best results for itself, and so there is an incentive for an individual firm to break the cartel agreement by secretly increasing its output and selling it at the fixed cartel price.

However, if all firms increased their output in this way, the cartel would collapse because the high price could not be sustained without a restricted output, and excess supply on the market would force down the price.

This has been the experience in recent years amongst the oil-producing countries of OPEC (the Organisation of Petroleum Exporting Countries). Attempts to agree on a restricted output quota for each country in order to push up oil prices have often broken down because some member countries exceeded their quota, or sold below the cartel's agreed price.

The success of a price cartel will depend on:

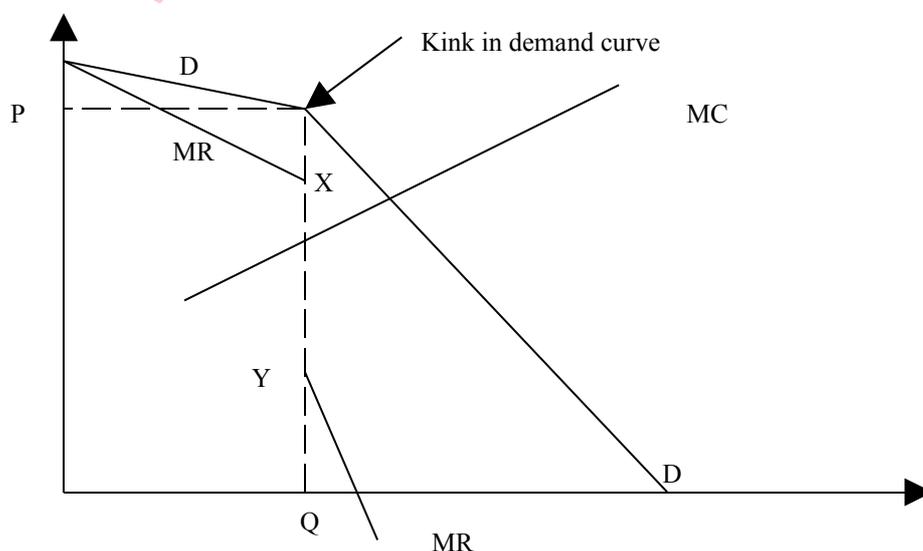
- (a) whether it consists of most or all of the producers of the product;
- (b) whether or not there are close substitutes for the product. For example, a price cartel by taxi drivers might lead to a shift in demand for transport services to buses, cars and trains;
- (c) the ease with which supply can be regulated. (In the case of primary commodities, such as wheat, rice, tea and coffee, total supply is dependent on weather conditions and even political events in the producing country);
- (d) the price elasticity of demand for the product. An attempt to raise prices might result in such a large a fall in demand that the total income of producers also falls (i.e. price elasticity is greater than 1);
- (e) whether producers can agree on their individual shares of the total restricted supply to the market. This is often the greatest difficulty of all.

The kinked oligopoly demand curve

Price cartels, whether 'official' or tacit and collusive, do not always exist in an oligopoly market. So how does an oligopoly firm which is competing with a rival oligopoly firm decide on its price and output level?

A feature of oligopoly markets, remember, is that each firm's pricing and output decisions are influenced by what the firm's rivals might do.

When demand conditions are stable, the major problem confronting an oligopolist in fixing his price and output is judging the response of his competitor(s) to the prices he has set. An oligopolist is faced with a downward sloping demand curve, but the nature of the demand curve is dependent on the reactions of his rivals. Any change in price will invite a competitive response. This situation is described by the kinked oligopoly demand curve in Figure 4, in which the oligopolist is currently charging price P , for output OQ , which is at the kink on the demand curve DD .



The kinked demand curve is used to explain how an oligopolist might have to accept price stability in the market.

If the oligopolist were to raise his prices above P , his competitors would keep their price lower and so many consumers would buy from them instead. An example is the difficulty which individual petrol companies have in raising the price of petrol at garages. If competitors do not raise their prices too, the firm usually soon has to restore its prices to their previous level. The demand curve would therefore be quite elastic at these higher prices.

If, on the other hand, the oligopolist were to reduce his prices below P , competitors would probably do the same. Total market demand might rise, but the increase in demand for the oligopolist's products would probably be quite low, unless demand in the market as a whole is elastic. The elasticity of demand at prices below p will be less than the elasticity at prices above p (hence the kink in the demand curve).

The MR curve is discontinuous at the output level where there is the kink in the demand curve. The kink in the demand curve explains the nature of the marginal revenue curve MR. At price P , output OQ , the MR curve falls vertically because at higher prices the MR curve corresponds to the more elastic demand curve, and at prices below p the MR curve corresponds to the less elastic demand. A firm maximises its profit at the point where $MR = MC$. There is a strong probability that the MC curve will cut the MR curve somewhere between points X and y (the discontinuous part of the MR curve). The more inelastic the demand curve is below price P , the longer the discontinuous portion (XY) of the MR curve will be.

The oligopolist's cost structure can change, with worsening or improved efficiencies, but as long as the MC curve cuts the MR curve through its vertical portion XY , the oligopolist's price and output decision should not alter. Hence, there will be price and output stability, with cost changes for the oligopoly firm, which change its MC curve, not affecting output and price.

Only if marginal costs rise far enough for the MC curve to pass through the MR curve above point X in Figure 4 is there a case for raising price, and only if MC falls far enough to pass through the MR curve below point y is there a case for lowering price.

In general, oligopoly prices will rise only if all the firms follow the lead of a rival in raising its price, so that the AR curve shifts outwards. The kink rises to the new common price level, which is again stable. The converse holds for price falls, perhaps occurring because of technological advance.

Price leadership and price wars in oligopoly markets

In oligopoly markets there is a tendency for one firm to set the general industry price, with the other firms following suit. This is called price leadership. It is one source of stability in a market where there may be cartels which tend to be undercut, and price wars.

When demand conditions change, the situation becomes somewhat different and price stability might no longer exist.

- (a) If total market demand falls, oligopolists might try to increase their share of the market by cutting prices.
- (b) Similarly, if one oligopolist begins to lose his share of the market, he might try to restore it by cutting prices. The consequence would be a price war. In the UK in recent years there have been price wars by supermarkets and oil companies, for example. The effect of price wars is usually beneficial to consumers, but they are of limited duration because it is not in the interests of oligopolists to sustain them for long.

Oligopoly: conclusion

We can conclude that oligopoly will lead to equilibrium prices and output in between the levels that would arise from monopoly and perfect competition, unless there is collusion, in which case there will be the same results as in monopoly.

Money and Banking

Your objectives

- (a) understand the nature and function of money
- (b) be aware of the main monetary aggregates and how they are made up
- (c) understand the role of banks in creating credit

Money and its functions

Money is used as a means of paying for goods and services, and paying for labour, capital and other resources. Money is important in an economy because:

- (a) it 'oils the wheels' of economic activity, by providing an easy method for exchanging goods and services (i.e. buying and selling);
- (b) the total amount of money in a national economy may have a significant influence on economic activity and inflation.

In this chapter, we shall consider:

- (a) the functions of money;
- (b) different means of quantifying the amount of money in an economy; and
- (c) how money is lent and borrowed, to move it from people who have more money than they can use to people who want to use it.

The functions of money

Attempts to define money have traditionally started with identifying what money does, ie what are the functions of money? We can identify four different functions of money.

Money acts as:

- (a) a means of exchange;
- (b) a unit of account;
- (c) a standard of deferred payment;
- (d) a store of value.

Money as a means of exchange

Without money, the only way of exchanging goods and services would be by means of barter, ie by a direct exchange of goods or services. If a shoemaker wanted to buy a horse, he would have to either:

- (a) find a horse-owner prepared to exchange a horse for a sufficient quantity of shoes of equal value to the horse; or else

- (b) find other people willing to exchange different goods (e.g. food, clothes etc) for shoes, and then trade these goods in exchange for a horse from the horse-owner.

A monetary economy is the alternative to a barter economy, and it provides a means of encouraging economic development and growth.

- (a) People are prepared to organise and work for an employer, and in return receive money wages.
- (b) A business will exchange its goods or services for money in return.
- (c) People will payout money in order to obtain goods or services.

Money as a unit of account

This function of money is associated with the use of money as a means of exchange. Money should be able to measure exactly what something is worth. It should provide an agreed standard measure by which the value of different goods and services can be compared.

For example, suppose that only four products are traded in a market: pigs, sheep, hens and corn. The relative value of these products must be agreed before exchange can take place in the market. It might be decided that:

- (a) 1 pig has the same value as 0.75 sheep, 3 hens or 1.5 bags of corn;
- (b) 1 sheep is the same value as 1.33 pigs, 4 hens or 2 bags of corn;
- (c) 1 hen is worth 0.33 pigs, 0.25 sheep or 0.5 bags of corn;
- (d) 1 bag of corn has the same value as 0.67 pigs, 0.5 sheep or 2 hens.

The function of money in the economy would be to establish a common unit of 'value measurement' or 'account' by which the relative exchange values or prices of goods can be established.

Money as a standard of deferred payment

When a person buys a good or service, he might not want to pay for it straightaway, perhaps because he has not yet got the money. Instead, he might ask for credit. Selling goods on credit is not an essential feature of an economy, but it certainly helps to stimulate trade. The function of money in this respect is to establish, by agreement between buyer and seller, how much value will be given in return at some future date for goods provided/received now.

Similarly, when a buyer and seller agree now to make a contract for the supply of certain goods in the future, the function of money is to establish the value of the contract, ie how much the buyer will eventually pay the seller for the goods.

In order to provide an acceptable standard for deferred payments, it is important that money should maintain its value over a period of time.

Suppose, for example, that a customer buys goods for an agreed sum of money, but on three months' credit. Now if the value of money falls in the three-month credit period, the sum of money which the seller eventually receives will be worth less than it was at the time of sale. The seller will have lost value by allowing the credit.

When the value of money falls (or rises) over time, sellers (or buyers) will be reluctant to arrange credit, or to agree the price for future contracts. Money would then be failing to fulfil its function as a standard for deferred payments.

Money can lose value as the result of price inflation. When inflation is high:

- (a) sellers will be reluctant to allow credit to buyers. For example, if a buyer asks for three months' credit, and inflation is running at 20% per annum, the 'real' value of the debt that the buyer owes will fall by about 5% over the three month credit period;
- (b) sellers will be reluctant to agree to a fixed price for long-term contracts. For example, a house-builder might refuse to quote a price for building a house over a twelve month period, and instead insist on asking a price which is 'index-linked' and rises in step with the general rate of inflation.

Money as a store of value

Money acts as a store of value, or wealth. So too do many other assets (eg land, buildings, art treasures, motorcars, machinery) some of which maintain or increase their money value over time, and some of which depreciate in value.

This means of course that money is not the only asset which acts as a store of wealth, and we need to extend our definition of this function of money.

Money is more properly described as acting as a liquid store of value. This definition has two parts to it.

- (a) Money is a store of value or wealth. A person can hold money in the certainty that its value does not fall and that it will have the same exchange value in the future that it does now, in 'normal' terms at least (i.e. ignoring inflation).
- (b) Money is a liquid asset.

The erosion of the value of money due to inflation provides one good reason why someone with wealth to store should hold assets which are not money.

Liquidity

A liquid store of value means that the wealth can be converted immediately (or at least very quickly) into a means of exchange for obtaining goods or services. Liquidity is therefore 'the ability to transform wealth holding into any form without loss of face value or delay'.

Liquidity is sometimes defined as 'readily convertible into cash', and the most liquid asset of all is cash itself.

There are two parts to this definition of liquidity. A liquid asset is one which can be converted into cash (or into a means of exchange for goods or services):

- (a) without significant delay; and
- (b) without significant penalty or loss of face value.

Let's look at each part of the definition in turn.

Delay, Liquidity is the ability to transform an asset into a means of exchange with minimum delay.

A liquid store of wealth can therefore be drawn on by its owner to obtain goods and services whenever he wants, and without having to wait to convert the store of wealth into a means of exchange. A painting is a store of wealth, but if its owner wishes to use this wealth to buy something else, he must first sell the painting and then use the proceeds from the sale to make his purchase. Since selling a painting takes time, a painting is not a liquid asset (and nor is a house).

Loss of face value, or significant penalty. A non-cash asset is liquid if it can be converted into cash within a short period of time without significant penalty. Significant penalty means:

- (a) loss of capital value, or face value; or

- (b) loss or forfeit of a substantial amount of interest, (say, loss of 14 days' interest or more).

Liquidity and interest

An alternative to holding money as banknotes, or money in a non-interest-bearing current bank account, is to hold on to an interest-bearing financial asset. Examples of such assets are:

- (a) a time deposit with a bank, or an interest-bearing current account;
- (b) a bond or debenture

These provide the holder with:

- (a) an interest yield on the value of the asset;
- (b) the certainty of a money repayment to the holder of the asset at the end of a certain period of time or after a certain period of notice has been given by the asset holder.

The holder of a financial asset sacrifices some liquidity in exchange for an interest yield on his store of wealth.

Narrow money and broad money

It is not always easy to decide whether a particular financial asset (e.g. a bank deposit) is money or not. In defining money, we can distinguish between narrow money and broad money.

Narrow money. 'Narrow money' refers to financial assets (including cash) which perform the functions of money when a fairly narrow definition of liquidity is applied. Financial assets must have a high degree of liquidity to be regarded as narrow money.

A definition of narrow money is 'money balances which are readily available to finance current spending, that is to say for "transactions purposes"'.

Broad money. In contrast, broad money includes financial assets which are relatively liquid, but not as liquid as narrow money items. A financial asset which would be regarded as narrow money would also fall within the definition of broad money; but broad money, as its name implies, extends the range of assets which are regarded as money.

Broad money refers generally to 'money held for transactions purposes and money held as a form of saving. It provides an indicator of the private sector's holdings of relatively liquid assets -assets which could be converted with relative ease and without capital loss into spending on goods and services'.

Monetary aggregates

Introduction

The money supply is the total amount of money in the economy. It is also referred to as the money stock. A monetary aggregate is a total of the money stock or money supply. In the UK, there are four different monetary aggregates which are published by the Bank of England. These are:

MO, M2, M4 and M3H,

Various other aggregates, including M1, M3 and M5 have been published at various times in the past. The main purpose of measuring a monetary aggregate is to discover by how much (and how rapidly) the money supply is rising in the economy, and

- (a) to predict from this rise what future changes in economic activity might be;

- (b) also to discover whether past changes in the money supply help to explain changes in economic activity which have already occurred.

There is also the view that by controlling the rate of increase in the money supply, inflation can be brought under control and economic conditions made more suitable for achieving economic growth and fuller employment.

Flows of funds within an economy

Introduction

The flow of funds describes the movement of funds or 'money' between one group of people or institutions in the economic system and other groups.

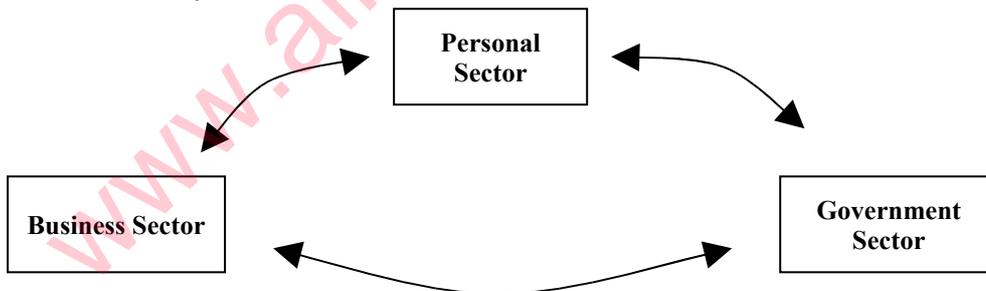
If we begin by ignoring imports and exports and foreign investments, we can start to build up a picture of the flow of funds by identifying three sectors of the economy:

- (a) the personal sector (i.e. individuals);
- (b) the business sector (or 'industrial and commercial sector'), i.e. companies and other businesses;
- (c) the government sector (i.e. central government, local government and public corporations).

Within each of these three sectors, there is a continual movement of funds. For example:

- (a) individuals will give money or lend money to other individuals;
- (b) companies will buy goods and services from other companies, and may occasionally lend money direct to other companies;
- (c) central government will provide funds for local government authorities and loss. making nationalised industries.

As well as movements of funds within each sector, there is also a flow of funds between different sectors of the economy.



But 'reality' is not quite so simple, and our analysis of the flow of funds in the UK has to take account of:

- (a) the overseas sector i.e. businesses, individuals and also governments in other countries who:
 - (i) sell goods or services to the UK;
 - (ii) buy goods or services from the UK;
 - (iii) invest capital in the UK; or
 - (iv) obtain investment capital from individuals or businesses in the UK;

(b) financial intermediaries.

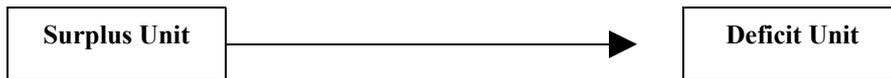
www.allonlinefree.com

Financial intermediaries

An intermediary is a go-between, and a financial intermediary is an institution which links lenders with borrowers, either (as principal) by obtaining deposits from lenders and then re-lending them to borrowers or (as broker) arranging a transaction. Here, we are concerned with the former type of intermediary.

The basic process of financial intermediation can be shown by simple diagrams.

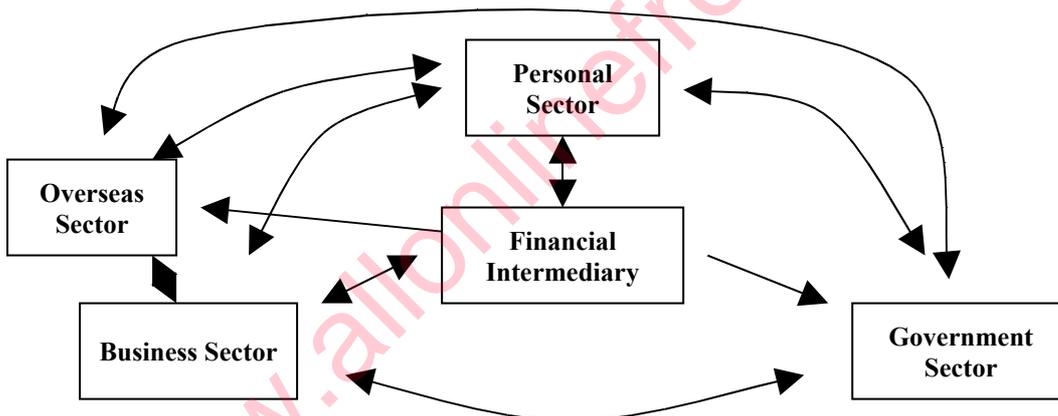
- (a) If no financial intermediation takes place, lending and borrowing will be direct.



- (b) If financial intermediation does take place, the situation will be:



Financial intermediaries might also lend abroad or borrow from abroad, and a final version of the diagram of the flow of funds is now as follows:



In the UK financial intermediaries include:

- (a) banks
- (b) building societies
- (c) insurance companies, pension funds, unit trust companies and the investment trust companies

The importance of banks as financial intermediaries

In spite of competition from building societies, insurance companies and other financial institutions, banks are still the major financial intermediaries.

- (a) The clearing banks dominate the retail banking market although competition from the building societies has been growing in the UK.
- (b) There is greater competition between different banks (overseas banks and the clearing banks especially) for business in the 'wholesale' lending market.

Risk reduction and maturity transformation

It is useful to understand what the benefits of financial intermediation are.

- (a) Financial intermediaries provide obvious and convenient ways in which a lender can save money. Instead of having to find a suitable borrower for his money, the lender can deposit his money with a bank, building society, pension fund, investment trust company, National Savings scheme etc. All the lender has to do is decide for how long he might want to lend the money, and what sort of return he requires, and he can then choose a financial intermediary that offers a financial instrument to suit his requirements.
- (b) Financial intermediaries aggregate (i.e. package up) the amounts lent by savers and lend on to borrowers in bigger amounts.
- (c) Provided that the financial intermediary is itself financially sound, the lender would not run any risk of losing his investment. Bad debts would be borne by the financial intermediary in its re-lending operations.
- (d) They provide a ready source of funds for borrowers. Even when money is in short supply, a borrower will usually find a financial intermediary prepared to lend some.
- (e) Financial intermediaries, most importantly, provide maturity transformation; in other words they bridge the gap between the wish of most lenders for liquidity and the desire of most borrowers for loans over longer periods. They do this by providing investors with financial instruments which are liquid enough for the investors' needs and by providing funds to borrowers in a different longer term form.

Banking

The banks and the banking system

Bank deposits are an important feature within most definitions of the money supply. There are different types of banks, and you will probably have come across a number of terms which describe them.

- (a) Clearing banks are the banks which operate the so-called 'clearing system' for settling payments (e.g. payments by cheque by bank customers).
- (b) The term 'retail' banks is used to describe the traditional High Street banks. The term 'wholesale' banks refers to banks which specialise in lending in large quantities to major customers. The clearing banks are involved in both retail and wholesale banking but are commonly regarded as the main 'retail' banks.
- (c) Merchant banks are banks which offer services, often of a specialised nature, to corporate customers.
- (d) Commercial banks refer to any banks which make commercial banking transactions with customers. They are distinct from the country's central bank.

Credit and banking

Before we go on to consider the role of banks in creating credit, it is worth giving some attention to the importance of credit in the economy.

Credit means owing money or being owed money

- (a) Taking or obtaining credit involves an arrangement whereby someone either:
 - (i) borrows money, with an undertaking to pay it back in time with interest;

- (ii) buys goods or services without paying for them immediately, but with an undertaking to pay for them in the future. Firms often buy goods from each other under this sort of credit arrangement, which is then referred to as 'trade credit'.
- (b) Lending or giving credit is the other side of the transaction. The person or organisation giving credit is the 'creditor' and the borrower is the 'debtor'.

Our main interest in credit for the purpose of this chapter concerns lending and borrowing money, rather than with buying goods on trade credit.

The functions of credit

The functions of credit can be seen from the point of view either of the borrower or the lender .

- (a) For the borrower, the reason for borrowing money is to be able to purchase goods or services 'now' that he wouldn't otherwise be able to afford. The borrower wants to buy now and pay later .
- (b) For the lender, the reason for lending money is that there is nothing that he particularly wants to spend his money on, and by lending it, he can earn some interest.

The borrower can be described as a 'deficit unit' with not enough money to buy all the goods and services he wants. The lender can be described as a 'surplus unit' with more money at the moment than he needs to spend.

Credit involves the transfer of money from a surplus unit to a deficit unit, in return for a promise to pay interest.

Bank deposits and the creation of money

Banks are major providers of credit. Banks create money when they give credit, and so bank lending has two aspects which are important for the economy:

- (a) credit and therefore expenditure in the economy;
- (b) increases in the money supply.

When a bank lends money, most of the money will find its way back into the banking system as new customer deposits. This adds to the money supply if we include bank deposits within our definition of the money stock.

Suppose, for example, that in a country with a single bank, customer deposits total £100,000. The bank, we will assume, re-lends all these deposits to other customers. The customers will use the money they have borrowed to buy goods and services and they will pay various firms and individuals for these purchases. If the firms and individuals receiving payment then put the money into their own accounts with the bank, the bank's deposits will have doubled.

It is this fact that most additions to bank lending end up as money in someone's bank account, adding to total customer deposits with the banks, that give banks this special ability to create money. It is an ability that is also shared in the UK by building societies, since building society deposits are included in definitions of the money supply (e.g. M2 and M4).

Fractional reserve systems

A bank could, in theory, go on re-lending all the deposits it receives, and if all the money re-lent then finds its way back into bank accounts, there is theoretically no limit to the amount of extra money that banks can create. 'This seeming power of the banking system to create apparently unlimited quantities of money has troubled economists and banking practitioners for many years. They have held up the spectre of an uncontrolled banking system creating money in such quantities that inflation would run riot and the value of money become debased.' (Crockett, Money, theory, policy and institutions)

This theoretical situation does not exist in practice, and there are two major reasons why this is so.

- (a) Banks do not re-lend all the extra deposits they receive. One obvious reason for this is that they need some liquidity to meet demands for withdrawals by customers. One way of ensuring adequate liquidity is to maintain a safe proportion of the bank's assets in the form of cash or liquid assets. If a bank does try to maintain a minimum ratio (fraction) of cash or liquid assets to total assets, it is said to be operating a fractional reserve system, or is said to be using a 'cash ratio' or 'reserve asset ratio'. A 'reserve asset' ratio is the ratio of a bank's reserve assets to its total assets. Reserve assets can be thought of as liquid assets or near-liquid assets, and so a reserve asset ratio is similar to a cash ratio.
- (b) Not all the money lent as bank loans to customers finds its way back into the banking system: there are 'leakages' (discussed later in this chapter).

A simplified illustration of a fractional reserve system cannot show the complexity of the financial system and the wide range of liquid and near-liquid assets a bank can invest in. In the illustration which follows, simplifying assumptions have had to be made for the sake of clarity.

Minimum cash ratio system illustrated: credit creation

We shall assume for simplicity that there is only one bank in the banking system, and that all money lent by the bank is re-deposited by various customers.

A customer depositing £1,000 in cash with the bank creates an asset for himself, in the sense that the bank, in return for the deposit, gives the customer an IOU, i.e. a promise to pay on demand, or subject to notice, the £1,000 deposited. The IOU is, reflected in the account opened under the name of the client. To the bank, the deposit is a liability. However, the deposit provides funds for the bank to acquire assets. We shall begin by assuming that the bank holds these assets entirely in the form of cash.

If the bank keeps the full £ 1,000 and does nothing with it, then it would simply operate as a 'cloakroom' for the client and the money so deposited. However, if the bank believes that the client is unlikely to claim the full £1,000 for some time, there will be some incentive to use the money rather than to keep it idle. One possibility would be to lend it; the bank would be taking a risk that it will not have the cash when its customer wants to have it back, but at the same time it would expect to make a profit by charging interest on the sum of money so lent.

Thus there is the liquidity /security /profitability conflict in the motives of the bank. On one hand the deposit of the £1 ,000 creates the opportunity for the bank to make a profit in the form of the interest that it can charge on the money it lends (the incentive being all the stronger if the bank is paying interest on the deposits it accepts), but on the other hand there is a risk that when the money is out on loan the client may claim it back. The bank will then be unable to meet its obligation to repay the cash to the client unless it can recall the loan instantly, which is unlikely.

As long as the bank feels that the likelihood is small that its depositors will demand a substantial proportion of their deposits in cash, then it faces an acceptable risk in lending some of the money. In other words, there is a balancing act between the desire to play it safe and keep the cash and the desire to make profits by lending. In practice, banks can be confident that in normal times, provided that there is not a serious loss in confidence in the banking system, depositors overall will not demand substantial proportions of their deposits as cash.

In the example in the table below, it is assumed that the bank has decided on the basis of past experience and observation to keep 50 pence in cash for every £1 deposited, and then lend out the other 50 pence. In other words, the bank in this example is operating a 50% cash ratio.

<i>Bank's liabilities</i> (= customer deposits)	<i>Bank's assets</i> (= cash or loans to customers)
(1) £1,000 deposits	£1,000 cash

The bank has £1,000 in cash. This is enough to support total assets of £2,000 and maintain a 50% cash ratio. If the bank now lends £1,000, and if all of that £1,000 is then spent by the borrowers but ends up back with the bank as deposits of other customers, the new situation for the bank will be as in (2) below.

(2)	£2,000 deposits	£1,000 cash
		£1,000 loans

Money creation and the cash ratio: credit multiplier

If the bank decided that the 50% cash ratio was too conservative and reduced it to 25% then for £1,000 cash deposited with the bank deposits could be expanded fourfold. It is important to understand that banks in the process of lending are also potentially creating money because clients either borrowing or receiving the proceeds of borrowers' expenditure can use their deposits to make money transactions. In a modern economy a large proportion of money transactions are by cheque or other bank transfer such as direct debits or standing order, thus transferring bank deposits from person to person or from firm to firm. Consumers and producers thus use the liabilities of a private institution (which is what bank deposits are) as money.

The fact that banks do not need to keep a 100% cash ratio automatically implies that they have the capacity to create money out of nothing. The size of this credit expansion depends on the size of their cash ratio.

We can summarise the quantitative side of credit creation in banks by writing the credit multiplier as:

$$\text{Deposits} = \frac{\text{cash}}{\text{cash ratio}} \quad \text{or} \quad D = \frac{c}{r}$$

The smaller the cash ratio the bigger the size of the deposits that a given amount of cash will be able to support.

This theoretical description of the credit multiplier applies to some extent in practice but there are other factors which operate to restrict the volume of expansion in bank deposits. In other words, if a bank decides to keep a cash reserve ratio of 10%, and it receives additional deposits of £1,000, the total increase in bank deposits will not be £1,000 + 10% = £10,000, but considerably less than this.

The constraints on the growth of a bank's deposits are:

- (a) leakages of cash out of the banking system;
- (b) the nature of customer demand for loans. Customers might not want to borrow at the interest rates the bank would charge;
- (c) prudent management of lending operations by the banks themselves. Banks should not lend to high-risk customers without good security.

Leakages, reducing the size of the credit multiplier

Leakages are caused by a few factors including the decision by the public (firms and individuals) to hold some of their extra money in cash (i.e. notes and coin in their hands). For example, if a bank lends £10,000 to a customer, who then uses the money to buy a car, the car dealer who is then paid the £10,000 might decide to hold some of his extra money in cash instead of putting it all into the bank. If we assume that the total amount of cash held by the public increases or decreases as the total volume of bank deposits rises or falls, it can be shown mathematically that the effect of the credit multiplier on total bank deposits is

$$D = \frac{C}{c+r}$$

where

D	is the eventual total increase in bank deposits
C	is the initial increase in bank deposits
r	is the banks' cash reserve ratio
c	is the ratio of cash held by the public to the volume of bank deposits.

For example, if the banks aim to have a cash reserve ratio of 10% and the public chooses to have a cash: bank deposits ratio of one-sixteenth (6.67%), then the total increase in bank deposits arising from an initial increase of £1,000 would be:

$$\frac{\pounds 1,000}{0.0667 + 0.10} = \pounds 6,000$$

The credit multiplier is now only times 6, not times 10.

Credit control and government policy

If the government is worried about growth in the money supply, or inflation in the economy, it might consider the growth in credit to be a major cause of the problem.

This might suggest that all a government has to do to control inflation and growth in credit spending is to stop banks from lending too much, ie to place controls over bank lending.

Credit control policies have been attempted in the past in the UK, without any long-term success. The main problems in applying direct controls over bank lending are as follows.

- (a) Enforcing the controls could be difficult. Financial institutions might try to 'bend' or avoid the rules, in order to continue lending. There might also be more direct lending between investor and borrower.
- (b) The development of a strong capital market in the UK will be damaged by government restrictions on lending, given the liberalisation of capital markets around the world in recent years.
- (c) Restrictions on certain forms of lending (e.g. investment loans to small companies, or mortgage lending to house buyers) could have harmful economic or social consequences.

The building societies

The building societies developed in the UK as mutual organisations to enable members to save or to borrow money to finance house purchase. Building societies are now allowed by the Building Societies Act 1986 to provide certain banking services as well, such as personal lending to customers for purposes other than house buying. Building societies now offer cheque book accounts, cash cards and many other facilities that compete directly with the 'retail' banks.

The growing similarity between banks and building societies is further evident in:

- (a) the inclusion of building society deposits in the broad monetary aggregates M4 (as well as M2)
- (b) the ability of building societies to become public limited companies and, in effect, banks. The Abbey National was the first UK building society to do this in 1989.

The previous description of banks, and the ability of banks to create money by giving credit, therefore applies to a large extent to building societies too. However, building societies are still mainly engaged in mortgage lending for house purchase, and are not yet fully comparable with banks.

The Central Bank And The Financial Markets

The functions of the Bank of England

Introduction

A central bank is a bank which acts on behalf of the government to supervise, regulate and control the country's banking system. In the UK, the central bank is the Bank of England. The Bank of England is a nationalised corporation and is the central bank of the UK. It is run by a Court of Directors, consisting of the Governor, Deputy Governor, and some Executive Directors and part-time Directors.

The Bank of England, as the UK's central bank, has a variety of functions.

- (a) It acts as banker to central government and holds the 'public deposits'. Public deposits include the National Loans Fund, the Consolidated Fund and the account of the Paymaster General, which in turn includes the Exchange Equalisation Account.
- (b) It is the central note-issuing authority in the UK; it is responsible for issuing bank notes.
- (c) It is the manager of the National Debt (ie it deals with long-term and short-term borrowing by central government and the repayment of central government debt).
- (d) It is the manager of the Exchange Equalisation Account (ie the UK's foreign currency reserves).
- (e) It acts as the administrator of any exchange control regulations that might be in force.
- (f) It acts as advisor to the government on monetary economic policy.
- (g) It acts as agent for the government in carrying out its monetary policies.
- (h) It is the supervisor of the banking system.

It acts as the informal supervisor of other financial institutions such as the Stock Exchange.

In general terms, it is the central bank's function to co-operate with the central government in carrying out the government's monetary policy.

In the UK, the Bank of England is also:

- (a) a banker to the commercial banks. Each of the commercial banks keeps accounts with the Bank of England;
- (b) a lender to the banking system. When the banking system is short of money, the Bank of England will provide the money the banks need at an appropriate rate of interest. As we shall see, it does this mainly by purchasing 'eligible bills' from the discount market, and it might very occasionally lend direct to the discount houses as a lender of last resort. (The discount houses are special banks which have the privileged role of acting as go-betweens between the Bank of England and all the other commercial banks in the UK.) This aspect of the Bank's activities is very important, because it provides the means by which the government can influence or control the level of interest rates.

The Bank of England as a banker

The Bank's customers are:

- (a) the Government;
- (b) other banks in the financial system, and in particular the clearing banks and discount houses.

As already mentioned, each commercial bank has a bank account with the Bank of England. Banks use these deposits to settle debts between each other. They also use these deposits to pay amounts due to the government, and when the government makes payments to the banks, there will be a transfer of funds from the government's account with the Bank to the banks' accounts with the Bank.

A feature of the Bank's role as a banker is therefore:

- (a) daily transfers of funds between the deposits of the banks as inter-bank debts are settled;
- (b) daily transfers of funds between the government's public deposits and the bankers' deposits.

The Bank of England as the central note-issuing authority in the UK

The Bank of England is the central authority in the UK for the issue of banknotes. The profits from note issues are paid to the Exchequer. These profits arise from the fact that notes are interest-free, but the funds obtained by the Bank from the commercial banks as payment for the issue of notes are used to acquire interest-bearing securities.

The Bank of England as manager of the National Debt

The Bank of England issues government securities to raise funds on behalf of the government, and is active in the gilt-edged market and money markets, buying and selling government (and other) securities. The Bank also administers the repayment of government debt when these debts reach their maturity (redemption date).

Long-term government debt instruments are referred to collectively as gilt-edged securities or gilts. Gilts are interest-bearing securities which are issued by the government and bought by investors. Short-term government debt takes the form of Treasury bills, which are issued by the government and bought largely by the discount houses.

The Bank of England as manager of the Exchange Equalisation Account

This account represents the deposits of the nation's gold and foreign currency reserves. The Exchange Equalisation Account can be used to influence the exchange rate of sterling against other foreign currencies. The exchange value of sterling is allowed to float (ie rise or fall in value) against other currencies in response to market supply and demand, but when the government considers that sterling's value is too high or too low, the Bank will intervene.

The Bank of England is thus a major active participator in three markets:

- (a) the gilt-edged market;
- (b) the money market (more specifically, the discount market); and
- (c) the foreign exchange market.

The Bank of England's role as administrator of exchange controls and other duties

Exchange controls are restrictions on the flow of money into or out of a country. At the moment, there are no such controls at all in force in the UK because they were removed by the government in 1979.

However, should they ever be re-introduced, the Bank would have the responsibility for administering them. Similarly, the central bank of any country where exchange control regulations are in force would have the responsibility of enforcing them.

The important function of the Bank as administrator or agent for the government's monetary policy will be described later.

The Bank has the responsibility under the Banking Act for the supervision of the banking system. It must try to ensure that individual banks retain sufficient liquidity and do not undertake ventures that are too risky.

The Bank of England as lender of last resort to the commercial banking system

In the UK, the discount market provides a link between the banking system and the government (Bank of England) whereby the Bank of England lends money to the banking system, when banks which need cash cannot get it from anywhere else (e.g. from other banks).

- (a) The Bank will supply cash to the banking system on days when the banks have a cash shortage. It does this by buying eligible bills from the discount houses in exchange for 'cash'. Eligible bills are bills (i.e. debt instruments), which the Bank will be prepared to buy from the discount houses. Most of these are bills that have been issued by banks, which are on the Bank's list of 'eligible' banks.
- (b) The Bank will remove excess cash from the banking system on days when the banks have a cash surplus. It does this by selling bills to the discount houses, so that the discount houses obtain interest-bearing bills in place of the cash that they do not want.

The process whereby this is done currently is known as open market operations. This simply describes the buying and selling of eligible bills between the Bank and the discount market.

Open market operations and short-term interest rates

One aspect of open market operations is that they enable the government, through the Bank of England, to influence short-term interest rates.

When bills are bought and sold, they are traded at a discount to their face value, and there is an implied interest rate in the rate of discount obtained. Interest rates on bills traded in open market operations have an immediate influence on other money market interest rates, such as the London Inter-Bank Offered Rate (LIBOR), and these in turn influence the banks' base rates.

Thus, if interest rates in the discount market's open market operations went up, there would very soon be an increase in other money market rates, and then the banks' base rates, and so lending rates on bank loans and overdrafts.

Open market operations are an important feature of the government's monetary policy, which the Bank implements on its behalf.

You might sometimes hear on television or radio news that banks have lowered or increased their base rates in response to a 'signal' from the Bank of England. Such a signal comes from a change of interest rate at which the Bank of England will deal that day in its open market operations.

The relationship between the Bank of England and the discount houses

The history of the discount market dates back to the early nineteenth century. A very significant feature of the discount market, as it has developed, is the special relationship which exists between the Bank of England and the discount houses. This relationship still exists, although since October 1988, the Bank of England has been willing to grant a relationship with other institutions, similar to its relationship with the discount houses, provided that these institutions meet certain conditions for being considered 'acceptable' by the Bank.

The Bank will act as a lender of last resort to the banking system as a whole via the discount houses. In doing so, the Bank (and government) is able to use the discount houses as a go-between through which it can influence the amount of cash (and influence short-term interest rates) in the UK banking system.

The Bank's aim is two-fold:

- (a) to provide a mechanism whereby the Bank can regulate the cash position of the banking system;
- (b) and in doing so, to control or influence the level of short-term interest rates.

How does the Bank influence short-term interest rates?

By operating as a 'lender' of cash in the discount market, the Bank's stated aim is to influence very short-term interest rates, and so indirectly influence the general level of interest rates.

The discount houses must compete with each other to sell bills to the Bank when they are short of money, or buy bills from the Bank when they have a cash surplus. The discount houses must therefore select the interest rates (i.e. rates of discount) at which they wish to buy or sell, and offer this rate to the Bank. The Bank will say yes or no, i.e. it will agree to buy or sell, or it will refuse. (If it refuses, the discount houses can offer a new price.)

By rejecting offers, the Bank will influence money market interest rates. If the Bank is content with the pattern of rates implied by the offers, it is generally prepared to accept sufficient to balance the market. But if the rates which are offered conflict with the Bank's interest rate objective, all or part of the offers may be rejected.

Open market operations occur daily, although general large interest rate changes occur just occasionally.

Event

- (a) Bank of England buys eligible bills for cash. Interest rate implied in the purchase price discount.
- (b) Commercial banks, short of cash, draw on deposits with discount houses.
- (c) Cash enters the banking system from the Bank of England.
- (d) Discount houses compete with bids to sell eligible bills to the Bank of England.

The capital markets and the money markets

Introduction

A distinction is usually made between:

- (a) capital markets, which are financial markets for raising and investing long-term capital; and
- (b) money markets, which are financial markets for lending and borrowing short-term capital.

'Long-term' and 'short-term' are rather arbitrary distinctions. What do we mean by long-term and short-term capital?

- (a) By short-term capital, we mean capital that is lent or borrowed for a period which might range from as short as overnight up to about one year, and sometimes longer.
- (b) By long-term capital, we mean capital invested or lent and borrowed for a period of about five years or more, but sometimes shorter.

There is a 'grey area' between long-term and short-term capital, which is lending and borrowing for a period from about one to two years up to about five years, which is not surprisingly sometimes referred to as medium-term capital.

It is not just the term of the lending and borrowing that distinguishes the capital markets from the money markets. Another important difference is the financial instruments that are dealt in by each market.

- (a) The 'capital markets' is a term used to describe the institutions which exist to provide long-term and medium-term finance, and the 'financial instruments' -i.e. the forms that such finance takes.
- (b) These are distinct from the 'money markets' which are the markets for short-term borrowing, but where the amount lent and borrowed in one transaction will be large.

Firms obtain long-term or medium-term capital in one of the following ways:

- (a) as share capital
- (b) as loan capital.

The government borrows from a variety of sources, but the two major ways of capital borrowing by central government are:

- (a) issuing gilt-edged securities or 'gilts'. These are government loan stock, on which interest is paid until maturity of the loan, when the capital is then repaid in full;
- (b) National Savings. The National Savings Bank is a government institution set up to borrow on behalf of the government, mainly from the non-bank private sector of the economy. The National Savings Bank operates through Post Offices throughout the country.

Capital markets in the UK

There are several 'market places' for raising capital. In the UK there are:

- (a) the Stock Exchange. The Stock Exchange provides the 'main market' where:
 - (i) quoted companies (i.e. public limited companies whose shares are quoted on the Stock Exchange) can raise new funds by issuing new shares or loan stock;
 - (ii) investors can buy and sell 'second-hand' stocks and shares -i.e. stocks and shares that are already in issue;
- (b) the 'gilts' or gilt-edged market. This is the market for the government's long-term debt securities;
- (c) the Unlisted Securities Market or USM. This is a market where companies that are not big enough to obtain a full listing on the Stock Exchange can raise new capital by issuing shares. Like the Stock Exchange, the USM is also a market for 'second-hand' shares. The USM is regulated by the Stock Exchange. At the time of writing (August 1993) the future of the USM is uncertain;
- (d) banks. Banks can be approached directly by firms and individuals for medium-term and long-term loans as well as short-term loans or overdrafts;
- (e) building societies. This is a capital market where individuals can obtain capital to buy a home with a mortgage, as well as personal loans;
- (f) National Savings. This is a capital market where the government obtains capital by borrowing from private investors.

The contribution of the Stock Exchange to the economy

The Stock Exchange is an organised capital market which plays an important role in the functioning of the UK's capitalist economy. It is the main capital market in the UK.

- (a) It makes it easier for large firms and the government to raise long-term capital, by providing a 'market place' for borrowers and investors to come together.
- (b) The Stock Exchange publicises the prices of quoted (or 'listed') securities, which are then reported in daily national newspapers such as the Financial Times. Investors can therefore keep an eye on the value of their stocks and shares, and make buying and selling decisions accordingly.
- (c) The Stock Exchange tries to enforce certain rules of conduct for its listed firms and for operators in the market, so that investors have the assurance that companies whose shares are traded on the Exchange and traders who operate there are reputable. Confidence in the Stock Exchange will make investors more willing to put their money into stocks and shares.
- (d) The index of share prices on the Stock Exchange acts as an indicator of the state of the country's economy.

Share prices on the stock market

The price of shares on a stock market fluctuate up and down

- (a) The price of shares in a particular company which are not traded much might remain unchanged for quite a long time; alternatively, a company's share price might fluctuate continually throughout each day.
- (b) The general level of share prices, as measured by some form of index (e.g. in the UK, by the All-Share Index and the FT -SE 100 Index or 'Footsie' Index, which includes 100 leading shares), will go up or down each day.

Stock markets are nowadays 'international' in character, with major investors buying and selling shares on stock markets in different parts of the world (e.g. New York, Tokyo, Hong Kong, London). Share price movements therefore reflect international views on the performance of a particular country's economy and industries.

Lenders of capital: institutional investors as financial intermediaries

The lenders of capital include private individuals, who buy stocks and shares on the Stock Exchange, or who invest in National Savings or building societies. However, there are some important institutional investors, i.e. institutions which specialise in lending capital in order to make a return.

These institutions, which are financial intermediaries themselves, include the following:

- (a) pension funds. Pension funds invest the pension contributions of individuals who subscribe to a pension fund, and of organisations with a company pension fund;
- (b) insurance companies. These invest insurance premiums paid on insurance policies by policy holders. If you think about it, insurance companies, like pension funds must do something with the premiums they receive, and in practice, they invest the money to earn a return;
- (c) investment trust companies. These are companies whose business is investing in the stocks and shares of other companies and the government. In other words, they trade in investments;
- (d) unit trust companies. These are similar to investment trust companies, in the sense that they invest in stocks and shares of other companies. They then sell portions or 'units' of these investments to individual investors; venture capital organisations. These are organisations that specialise in raising funds for new business ventures, such as 'management buy-outs' (i.e. purchases of firms by their management staff). These organisations are therefore providing capital for fairly risky ventures. A venture capital organisation that has operated for many

years in the UK is 'Investors in Industry' or 'the 3i group'. In recent years, many more venture capital organisations have been set up, and there is an active British Venture Capital Association of such organisations.

The effectiveness of the capital markets as providers of capital to businesses

The functions of the capital market are to enable investors to lend money and borrowers to raise money, in such a way that new investments can take place which help to stimulate or add growth to the economy. Although the market exists and is active, it is perhaps doubtful whether it is as effective as it should be in providing a means of raising money for new investments.

- (a) In recent years, high rates of interest have made it difficult for companies to issue new loan stock. This may change now that interest rates have fallen and this kind of capital has become cheaper.
- (b) Public limited companies have been able to raise new funds from share issues on the Stock Market, but the Stock Market remains mainly a market for second-hand shares rather than new issues. Only public limited companies are permitted to issue shares in this way, and in spite of the USM, smaller firms have much more difficulty in finding investors who will subscribe for new shares. Venture capital organisations are not sufficiently large to supply the volume of equity finance to small firms that would probably be needed for an effective capital market
- (c) Several large firms experienced difficulties during the economic recession of the early 1980s and were taken into 'intensive care' by their bank. To this extent, the capital market functioned effectively to prevent an even deeper recession.
- (d) Bank lending and lending by other financial institutions are often accused of being excessively conservative, avoiding risky investments. Since economic growth will probably depend on the success of risky ventures, a conservative approach is likely to reduce the prospects of successful ventures in large numbers. However, the large number of receiverships in the recent recession suggest that this reputation of the banks for conservatism may not be wholly deserved. During the economic boom period of 1986 to 1988, the banks lent to numerous businesses which faced serious financial difficulties in the years that followed.

The money markets

The 'money markets' in the UK are operated by the banks and other financial institutions. Although the money markets largely involve wholesale borrowing and lending by banks, some large companies and the government are also involved in money market operations. The money markets consist of:

- (a) the discount market. This is a market where some banks (the discount houses) buy and sell bills of exchange. An important aspect of trading in bills is the daily trading in certain 'eligible bills' for cash between the discount houses and the Bank of England. The government uses this trading, known as open market operations, both to control or influence the level of short-term interest rates and also to ensure that the banking system as a whole has an adequate daily supply of cash to operate with;
- (b) the interbank market. This is the 'market' in which banks lend short-term funds to one another, and is a very important money market.

A key interest rate in the UK money market is the LIBOR, or London Inter-Bank Offer Rate, which is the rate of interest at which the largest London banks will lend to each other short-term. The level of the 3-month LIBOR influences clearing banks' base rates, and much 'big money' (or 'wholesale') lending of short-term funds to banks' customers is fixed at a certain margin above LIBOR (e.g. 1% above the 3 month LIBOR);

- (c) the eurocurrency market. This is the market operated by banks for lending and borrowing in foreign currencies. Most of the trading is done by banks. Firms wishing to borrow in a foreign currency will usually do so from a bank;
- (d) the Certificate of Deposit market. This is a market for trading in Certificates of Deposit;
- (e) the local authority market. This is a market in which local authorities borrow short-term funds from banks and other investors, by issuing and selling short-term 'debt instruments';
- (f) the finance house market. This refers to the short-term loans raised from the money markets by finance houses (e.g. hire purchase finance companies);
- (g) the inter-company market. This refers to direct short-term lending between companies, without any financial intermediary. This market is very small, and restricted to the treasury departments of large companies, and has largely been superseded by the sterling CP markets;
- (h) the commercial paper /MTN market. This is a market in which companies issue debt securities, known as commercial paper (CP) and medium-term notes (MTNs). The paper carries a rate of interest, and is purchased by investors.

A distinction is sometimes made between the discount market and all the other money markets (b) to (g) which are referred to collectively as the parallel markets or 'wholesale' markets.

NATIONAL INCOME ANALYSIS

Your objectives

After completing this chapter you should:

- (a) be aware of the different ways of measuring a nation's economic wealth;
- (b) understand the way in which income flows between households and firms;
- (c) understand how national income is measured.

Measuring the creation of economic wealth

Introduction

We start our analysis of issues in macroeconomics by looking at how we measure the total amount of economic activity in a country. Three key measures are:

- (a) national income;
- (b) gross national product (GNP);
- (c) gross domestic product (GDP).

These are related but different measures of the economic wealth that a country creates or earns over a period of time, usually one year.

Why is national income so important?

National income is an important measure because it is an aggregate of personal incomes. The bigger the national income in a country, the more income its individual inhabitants will be earning on average. More income means more spending on the output of firms, and more spending (ignoring inflation) means more output of goods and services.

National income provides information of great importance to governments in setting their macroeconomic objectives. Such objectives typically include:

- (a) full employment. In the UK in recent years employment has been at levels way above the three per cent which is sometimes argued to be the irreducible minimum for this country. Some commentators believe the reason for this to be the low priority accorded by the present Conservative government to full employment as a macroeconomic objective;
- (b) economic growth. This is an upward trend in the total output of a nation. Most governments will try to achieve a satisfactory rate of growth by adopting policies to encourage investment, research and development and so on;
- (c) stable prices. Rapid rates of price increases are thought to be undesirable for reasons discussed in a later chapter. Governments aim to maintain prices at a stable level or, if that is not possible (and it rarely is), at least to keep increases to a minimum;
- (d) a satisfactory balance of payments. This is discussed in detail in a later chapter.

Who creates economic wealth?

Economic wealth is created by:

- (a) the people or organisations that spend money to buy the goods and services:
 - (i) consumers (or 'households');
 - (ii) the government; and
 - (iii) foreign buyers (the 'overseas sector'); or
- (b) the people who earn the wealth:
 - (i) labour, who earn wages for the work they do;
 - (ii) providers of capital, who earn interest on the capital they invest;
 - (iv) owners of 'land', who earn rent on the land they provide for economic use;
 - (v) entrepreneurs, who earn profits for the business risks that they take.

As has often been stressed earlier in this book, the scarcity of these factors of production, and the need for decisions on their allocation, is the fundamental problem of economics;

- (c) the creators of wealth can also be identified as the firms (or government departments and corporations) which produce the goods or services in the national economy.

The three approaches (a), (b) and (c) give rise to three ways of defining economic wealth:

- (a) the expenditure approach;
- (b) the income approach; and
- (c) the output approach.

These are defined and discussed later in this chapter.

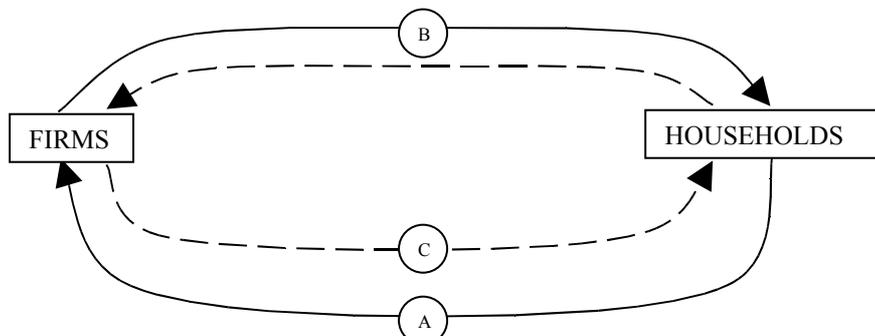
The circular flow of income

Introduction

Income in a country's economy flows between households and firms.

- (a) The income of firms is the sales revenue from the sales of goods and services.
- (b) The income of households is the income arising from the ownership of the factors of production.

Firms must pay households for the factors of production, and households must pay firms for goods. This creates a circular flow of income and expenditure, as illustrated in Figure I.



The total sales value of goods produced (output) should equal the total expenditure on goods, assuming that all goods that are produced are also sold. The amount of expenditure should also equal the total

income of households, because it is households that consume the goods and they must have income to afford to pay for them.

Households earn income. They earn income because they have provided the factors of production which enable firms to output goods and services. The income earned is used as expenditure on these goods and services that are made.

Withdrawals and injections into the circular flow of income

Our simplified diagram of the circular flow of income needs to be amended to allow for:

- (a) withdrawals from the circular flow of income. These are movements of funds out of the cycle of income and expenditure between firms and households;
- (b) injections into the circular flow of income. These are movements of funds in the other direction.

There are three types of withdrawal from the circular flow.

- (a) Savings (S). Households don't spend all their income. They save some, and these savings out of income are withdrawals from the circular flow of income quite simply because savings are not spent.
- (b) Taxation (T). Households must pay some of their income to the government, as taxation. Taxes cannot be spent by households, because the funds go to the government.
- (c) Imports (M). When we consider national income, we are interested in the economic wealth that a particular country is earning. Spending on imports is expenditure, but on goods made by firms in other countries. The payments for imports go to firms in other countries, for output created in other countries. Spending on imports therefore withdraws funds out of a country's circular flow of income.

There are also three types of injection into the circular flow of income.

- (a) Investment (I). Investment in capital goods is a form of spending on output, which is additional to expenditure by households. Just as savings are a withdrawal of funds, investment is an injection of funds into the circular flow of income, adding to the total economic wealth that is being created by the country.
- (b) Government spending (G). Government spending is also an injection into the circular flow of income. In most mixed economies, total spending by the government on goods and services represents a large proportion of total national expenditure. The funds to spend come from either taxation income or government borrowing.
- (c) Exports (X). Firms produce goods and services for export. Exports earn income from abroad, and therefore provide an injection into a country's circular flow of income.

Three approaches to measuring national income

As already mentioned, these are:

- (a) The expenditure approach. The economic wealth created in a period can be measured by the amount of expenditure on the goods and services that are produced by the nation's economy.
 - (i) The expenditures will be incurred by consumers, the government and foreign buyers (i.e. exports). Expenditures on imports represent wealth created by other countries, and so the value of expenditure on imports must be deducted from the total expenditure figure.
 - (ii) Expenditures by firms (e.g. companies) are excluded, to avoid double-counting.

- (b) The income approach. This approach measures the income of individuals from employment and from self-employment, the profits of firms and public corporations and rent on property. (Interest earnings will be included within the profits of companies or the income of individuals.)
- (c) The output approach. This approach is to measure the (sales) value of output produced by firms and other organisations in the period.

All three approaches will in theory result in the same total amount for economic wealth created in the period, which we call Gross Domestic Product or GDP. In practice, statistical discrepancies arise which cause differences between the alternative figures.

The government and national income

The government has several functions within the national economy, and so plays several different roles in the circular flow of income.

- (a) It acts as the producer of certain goods and services instead of privately-owned firms, and the 'production' of public administration services, education and health services, the police force, armed forces, fire brigade services, and public transport are all aspects of output. The government in this respect acts, like firms, as a producer, and must also pay wages to its employees.
- (b) It acts as the purchaser of final goods and services and adds to total 'consumption' expenditure. National and local government obtain funds from the firms or households of the economy (e.g. in the form of taxation) and then use these funds to buy goods and services from other firms.
- (c) It invests by purchasing capital goods (e.g. building roads, schools and hospitals).
- (d) It acts as a means of transferring wealth or income from one section of economy to another e.g. by taxing workers and paying pensions, unemployment benefits and social security benefits ('transfer payments') to other members of society.

Items (a), (b) and (c), but not (d), contribute to the creation of national income.

National income statistics

What is national income?

The UK national income can be defined as 'the sum of all incomes of residents in the UK which arise as a result of economic activity, that is from the production of goods and services. Such incomes, which include rent, employment income and profit, are known as factor incomes because they are earned by the so-called factors of production: land, labour and capital.' (CSO)

National income is also called net national product.

- (a) The terms 'income' and 'product' are just two different aspects of the same circular flow of income.
- (b) The term 'net' means 'after deducting an amount for capital consumption or depreciation of fixed assets'. We shall return to this point later.

Gross domestic product (GDP)

Most UK national income is derived from economic activity within the UK. Economic activity within the UK is referred to as total domestic income or domestic product. It is measured 'gross' (ie before deducting an amount for capital consumption or depreciation of fixed assets) and the term gross

domestic product therefore refers in the UK to the total value of income/production from economic activity within the UK.

Gross national product (GNP)

'Some national income arises from overseas investments while some of the income generated within the UK is earned by non-residents. The difference between these items is net property income from abroad.' (CSO)

Gross national income or gross national product (GNP) is therefore the gross domestic product (GDP) plus the net property income from abroad or after subtracting the net property income from abroad, if it is a negative value.

The relationship between GDP, GNP and national income

The relationship between GDP, GNP and national income is therefore as follows.

	GDP
<i>plus</i>	Net property income from abroad
<i>equals</i>	GNP
<i>minus</i>	Capital consumption
<i>equals</i>	National Income (net)

The expenditure approach to measuring national income

Probably the most widely used measure of national income is the measurement of total spending or expenditure and it is worth looking at this in some detail.

UK NATIONAL INCOME 1991: EXPENDITURE APPROACH

	£bn
<i>At current market prices</i>	
Consumers' expenditure	367.9
General government consumption	121.9
Gross domestic fixed capital formation	95.4
Value of increase/(decrease) in stocks and work in progress	<u>(5.3)</u>
Total domestic expenditure	579.9
Exports of goods and services	135.1
Imports of goods and services	<u>(140.4)</u>
Gross domestic product at market prices	574.6
Taxes on expenditure (indirect taxes)	<u>(83.0)</u>
	491.6
Subsidies	5.9
Statistical discrepancy	<u>(0.4)</u>
GDP -gross domestic product at current factor cost	497.1
Net property income from abroad	<u>0.3</u>
GNP -gross national product at current factor cost	<u>497.4</u>

Note the following points.

- (a) National income is calculated by subtracting a fairly arbitrary amount for depreciation from a total value known as Gross National Product, or GNP .
- (b) GNP itself is calculated by measuring total expenditure within the national economy (which we call Gross Domestic Product or GDP) and then adding on amount for income from property (assets) abroad.
- (c) The three measures, national income, GNP and GDP are therefore closely related. When we distinguish between the expenditure approach, the income approach and output approach, we are really talking about three approaches to calculating GDP .

- (d) We might refer to any of the three measures, national income, GNP or GDP. Where a government seeks to increase national income, it is also seeking increases in GDP and GNP.

Notes on the expenditure approach

Total spending then, consists of consumption spending, government spending, investment spending, spending by foreigners on our goods and services minus spending by us on foreign goods and services. This is often symbolised as $C + I + G + (X - M)$ where:

C	=	consumption expenditure
I	=	investment expenditure
G	=	government expenditure
X	=	expenditure on our exports by foreigners
M	=	expenditure by us on imports

From the table it can be seen that when we calculate $C + I + G + (X - M)$ the total we arrive at is called the gross domestic product at market prices. This measure (GDP at market prices) is one way of expressing the level of economic activity in the UK.

If a government is planning its economic policy, and wishes to increase the country's GDP and GNP, it might wish to turn its attention to any of these items, i.e.:

- (a) trying to increase consumer spending, C;
- (b) trying to increase private investment, I;
- (c) deciding to increase government spending, G and/or I;
- (d) trying to improve the balance of payments on overseas trade, $(X - M)$.

Since the prices of many goods and services are distorted by sales taxes (eg alcohol and cigarettes) and some are distorted by subsidies (e.g. many agricultural products) we often wish to view the situation without these distortions and convert GDP at market prices to GDP at factor cost:

- (a) expenditure at market prices, i.e. the actual amounts paid for the goods by their buyers; and
- (b) expenditure at factor cost, i.e. expenditure at market prices minus indirect taxes plus any government subsidies.

As you can see from the table this is still not the end of the story. Many statisticians, economists, governments and international agencies like to include property income from abroad to give a fuller picture of what is happening in the domestic economy. When this extra item is included we now have a measure called gross national product or GNP.

Although technically national income has a particular definition, generally you will find both measures (GDP and GNP) loosely referred to as 'national income'.

The income approach to measuring national income

The second method of calculating national income is the income method. Since money spent by an individual or firm must become income to another we should not be surprised to find that except for a residual error the results of the two methods are the same.

UK NATIONAL INCOME 1991: INCOME APPROACH

	£bn
<i>At current prices</i>	
Income from employment (wages and salaries, plus employers' NI contributions)	329.8
Gross trading profits of companies	60.7
Gross trading surplus of public corporations	3.1
Gross trading surplus of government enterprises	0.1
Other income	106.1
Less stock appreciation	<u>(2.8)</u>
GDP -income based	497.0
Statistical discrepancy	<u>0.1</u>
GDP -expenditure based	497.1
Net property income from abroad	<u>0.3</u>
GNP	<u>497.4</u>

More about the income based approach

When we refer to 'national income', it might seem more obvious to consider the income- based approach rather than an expenditure- based approach to measuring the statistics.

The table showing the income- based approach shows as separate items:

- (a) income from employment (i.e. wages and salaries before deducting tax and including employers' National Insurance contributions);
- (b) pre-tax profits of companies;
- (c) pre-tax profits of public corporations (including nationalised industries);
- (d) the pre-tax 'surplus' of other government enterprises.

Interest earned by individuals and companies on any investments they hold is included in the figures for (a) and (b).

You might notice that these income figures do not include:

- (a) income from government pensions or social security payments;
- (b) any value for work done by individuals for no monetary reward, such as housework done by housewives or do-it-yourself home improvements. These are activities for which no money value can be given, and so are not 'economic' activities.

Transfer payments (transfer incomes)

Transfer payments are payments that are made any contribution to output in return. They wealth, rather than a reward for creating

Transfer payments do not lead directly to any increase in marketable output of goods.

Most transfer payments are made by the government, which collects taxes and uses some of its tax income to make payments for:

- (a) old age pensions, and

(b) social security payments

Transfer payments are not included in GNP because they do not add to 'marketable' output of goods to someone, where the recipient does not make payments which involve the transfer of new economic wealth.

Government expenditure figures within GNP are net of transfer payments.

Rent

Rent is one component of 'other income' in the list of income earned (other components include income from self-employment). It is the profits (or 'operating surplus') from the ownership of land and buildings. It also includes 'a value for 'imputed rent' for owner-occupied homes. This is explained below.

Imputed income and expenditure

Although the income figures exclude most work that is done for no monetary reward (eg housework and do-it-yourself activities) the definition of production/economic activity in the UK does include some activities which are not carried out for a money reward.

The main one of these activities is the provision of owner-occupied houses. When a person buys his own home and lives in it, no money will be exchanged between the owner and the occupier because they are one and the same person. However 'the services of the house do nevertheless have a value equivalent to the net income which could have been obtained by letting it commercially. A figure based on this approach is included in the national income. In effect, the owner-occupier is divided into two separate transactors. It is supposed that as owner, he lets the house to himself as occupier for a certain rent.' (CSO)

This process of 'inventing a transaction' is known as imputation. In the case of owner-occupied dwellings an imputed amount of rent is included in the income statistics as rent (income-based approach) or consumers' expenditure (expenditure-based approach).

The output method of measuring national income

The third method of calculating national income is the output method. Since the goods and services we spend our money on must have been produced by some industry or another it is not surprising to find the amount we have all spent is the same as the value of the goods and services produced (except for a residual statistical discrepancy).

Estimates have to be made when accurate figures are unobtainable, and there are omissions in obtaining some figures, and deliberate errors in others -for example, fiddling tax returns might occur on a large scale. Because of errors, the three approaches will produce slightly different figures, and one of them must be taken as 'correct'. In practice, the expenditure based figures are considered most reliable, and the income based and output based GDP figures are adjusted to the expenditure based GDP figure, by inserting a balancing item known as a residual error.

National income and inflation

Inflation is a particular problem in using national income as a measure of national wealth. Price inflation increases the money value of national income.

We should be careful not to interpret this as meaning that there is more economic activity going on in our economy. All that has happened is that the prices of the things we are measuring have increased. To see if there has been any 'real' change in the level of activity we must deduct any influence due to inflation. Although this is not a simple operation, the standard method for turning 'money' GDP or GNP into 'real' measures is to use what is called the 'GDP deflator' in order to take inflation out of the figures.

If we are comparing one year's national income with another it is very important that we use 'real' figures if we are to reach any worthwhile conclusions, because with 'real' comparisons of national income from one year to the next, we are comparing like with like.

NATIONAL INCOME AND UNEMPLOYMENT

Your objectives

After completing this chapter you should:

- (a) understand the relationship between aggregate demand, aggregate supply and equilibrium national income;
- (b) understand the impact of consumption, savings and investment on national income;
- (c) understand the link between the level of national income and the level of employment.

Aggregate demand and supply and equilibrium national income

Introduction

You may well be aware that there has been some difference of opinion between economists as to how real economic growth can be achieved. Broadly speaking, economists can be divided into two camps, the 'Keynesians' and the 'monetarists'. These two camps have had differing ideas about how national income can be made to grow, how full employment can be achieved and how booms and slumps of trade cycles can be smoothed out. They differ in their views about the causes of inflation, the extent to which inflation creates unemployment and prevents economic growth, and the effectiveness of government measures to stimulate the economy, for example by borrowing and spending heavily to create new investment and jobs.

In this chapter we look at the basic elements of the Keynesian model for national income determination and equilibrium.

Keynesian economics originated with John Maynard Keynes, an English economist whose book *The General Theory of Employment, Interest and Money* (1936), revolutionised economic analysis. Keynes put forward his ideas following a period in which there was an economic boom (after the First World War), followed by the Wall Street Crash in 1929, and the Great Depression in the 1930s when unemployment levels soared.

Aggregate demand and aggregate supply

Keynes argued that the level of overall output and employment depends on the level of aggregate demand in the economy. His basic idea was that demand and supply analysis could be applied to macroeconomic activity as well as microeconomic activity.

- (a) Aggregate demand means the total demand in the economy for goods and services.
- (b) Aggregate supply means the total supply of goods and services in the economy.

Aggregate supply depends on physical production conditions -the availability and cost of factors of production and technical know-how. Keynes was concerned with short-run measures to affect the economy, and he also wrote in a period of high unemployment when there was obviously no constraint on the availability of factors of production. His analysis therefore concentrated on the demand side. 'Supply side' economics (discussed later) describes the views of economists who do not subscribe to the Keynesian approach to dealing with the problem of national income and employment, and prefer instead to concentrate on 'supply side' i.e. production factors).

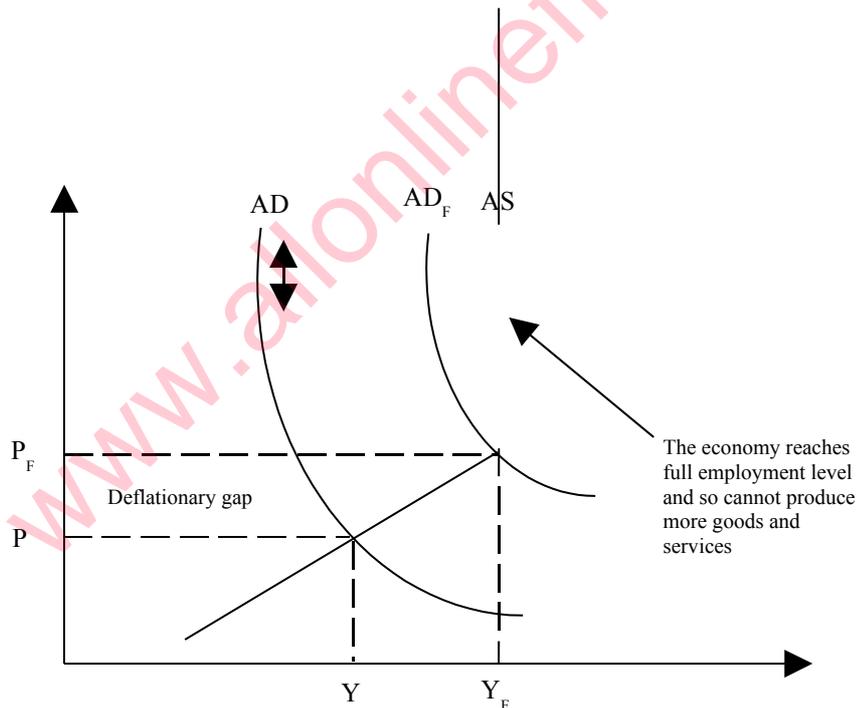
The aggregate supply curve will be upward sloping, for the reasons applying to the 'microeconomic' supply curves mentioned in earlier chapters. A higher price means that it is worthwhile for firms to hire more labour and produce more because of the higher revenue-earning capability. So at the macroeconomic level, an increasing price level implies that many firms will be receiving higher prices for their products and will increase their output.

In the economy as a whole, supply will at some point reach a labour constraint, when the entire labour force is employed. When there is full employment, and firms cannot find extra labour to hire, they cannot produce more even when prices rise, unless there is some technical progress in production methods. The aggregate supply curve will therefore rise vertically when the full employment level of output is reached (AS in Figure I).

Aggregate demand (AD) is total desired demand in the economy, for consumer goods and services, and also for capital goods, no matter whether the buyers are households, firms or government. Aggregate demand is a concept of fundamental importance in Keynesian economic analysis. Keynes believed that national economy could be 'managed' by taking measures to influence aggregate demand up or down.

Aggregate demand is the total desired demand, just as a 'microeconomic' demand curve represents the desired demand for a particular good at any price level. The AD curve will be downward-sloping because at higher prices total quantities demanded will be less.

Keynes argued that a national economy will reach equilibrium where the aggregate demand curve and aggregate supply curve intersect.



i.e., expanding either AD or AS in the economy. As suggested already, Keynesian economists concentrate on shifts in AD.

The Keynesian economic argument is that if a country's economy is going to move from one equilibrium to a different equilibrium, there needs to be a shift in the aggregate demand curve. To achieve equilibrium at the full employment level of national income, it may therefore be necessary to shift the AD curve to the right (upward) or the left (downwards).

Deflationary and inflationary gaps

In a situation where there is unemployment of resources there is said to be a deflationary gap (Figure 1). Prices are fairly constant and real output changes as aggregate demand varies. A deflationary gap can be described as:

'the extent to which the aggregate demand function will have to shift upward to produce the full-employment level of national income'.

(Lipsey: An introduction to positive economics)

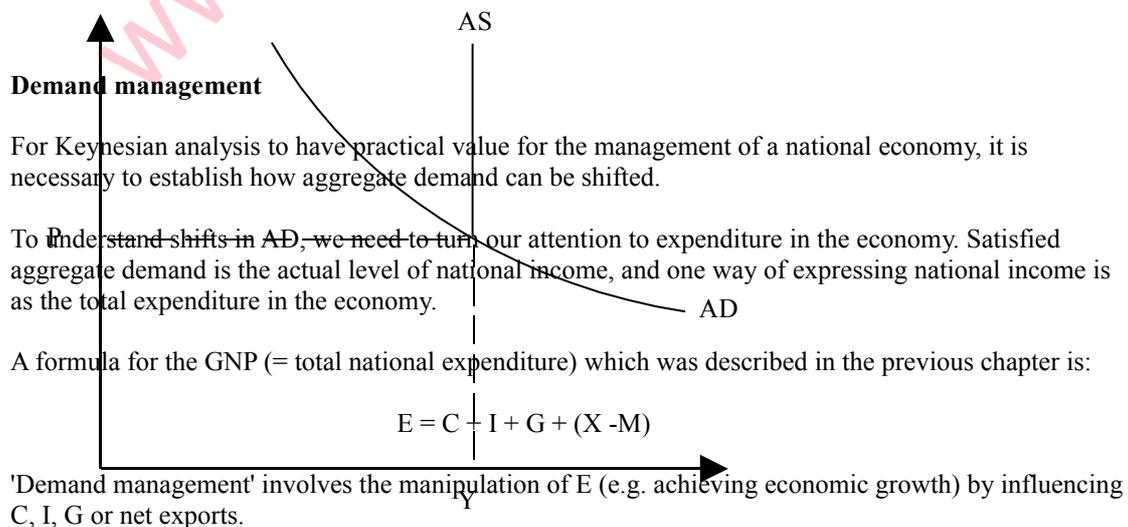
In a situation where resources are fully employed, there is said to be an inflationary gap, for changes in aggregate demand will cause price changes and not variations in real output. An inflationary gap can be described as:

'the extent to which the aggregate demand function would have to shift downward to produce the full employment level of national income without inflation '.

(Lipsey)

The 'ideal' equilibrium national income

If one aim of a country's economic policy is full employment, then the 'ideal' equilibrium level of national income will be where AD and AS are in balance at the full employment level of national income, without any inflationary gap i.e. where aggregate demand at current price levels is exactly sufficient to encourage firms to produce at an output capacity where the country's resources are fully employed. This is shown in Figure 2, where equilibrium output will be y (full employment level) and price levels P .



If we ignore capital consumption, we can equate E (GNP) with national income. This is what we shall do in our analysis of the Keynesian model.

Withdrawals and injections into the circular flow of income

For a national economy, there are certain withdrawals from and injections into this circular flow of income. Withdrawals divert funds out of the circular flow and injections add funds into it.

Keynes argued that for an equilibrium to be reached in the national income, not only must $AD = AS$, but also total planned withdrawals from the circular flow of funds must be equal to total planned injections. In other words, for equilibrium:

$$W = J \text{ and so } M + T + S = X + G + I.$$

In the long-term W will always equal J .

- (a) The difference between the value of imports M and the value of exports X is the balance of payments deficit. Even in the short-term, this difference must be balanced by borrowing or lending abroad, as we shall see in a later chapter.
- (b) The difference between government spending and taxation can only be made up by government borrowing. Loans are eventually repaid.
- (c) In the long-run, savings will also equal investments, even though the people who save and the firms who invest are not the same 'people'. We shall look more closely at savings and investment later .

However, although W and J will be equal retrospectively and in the long run, it does not follow that planned J and planned W will equal each other in the short run, since injections and withdrawals are made by different people.

It is precisely this 'frustration of plans' in the short run that causes national income to change over time. The imbalance between J and W creates factors which can make the level of national income change. Keynes argued that the imbalance between planned withdrawals and planned injections explained trade cycles, i.e. booms and slumps in the economy which prevent an economy from settling down at an equilibrium level.

Consumption, savings and investment

Consumption and savings (C and S)

Let us now go into a bit more detail of Keynesian analysis, and concentrate particularly on consumption, savings and investment. To simplify our analysis, we shall ignore government spending, taxation, imports and exports for the time being. By ignoring imports and exports, we are concentrating on a closed economy. i.e. a national economy which is not in any way dependent on foreign trade.

If we ignore G , T , X , and M , we can look at a circular flow of income in which households divide all their income between two uses:

- (a) consumption on goods and service; or
- (b) saving.

Provided that national income is in equilibrium, we will have:

$$Y = C + S \text{ (i.e. } Y \text{ is the same as } C + S \text{)}$$

where Y = national income, C = consumption and S = saving

This should seem logical to you. Income can only be either spent or saved. Since we have a closed economy, consumption must be of goods produced by the economy itself.

Savings

What people do not spend out of their income, they will save.

There are two ways of 'saving'. One is to hold the income as money (banknotes and coin, or in a current bank account). The other way is put money into some form of interest-bearing 'investment'.

In the long run, there is no reason for people to hold banknotes or keep money in a current bank account, unless they intend to spend it fairly soon. If this is so, income that is not spent will be saved and income that is saved will, eventually, be invested.

We can therefore conclude that in conditions of equilibrium for national income:

$$\begin{array}{lcl} Y & \equiv & C+S \quad \text{and} \\ Y & \equiv & C+I \quad \text{and so} \\ I & \equiv & S \end{array}$$

In the short-run, however, savings and investment might not be equal and so there might not be equilibrium.

The propensities to consume and save

In aggregate, the population will spend a certain proportion of total income on consumption. This proportion is known as the average propensity to consume (APC).

If there is an increase in total income some of the extra income will be spent and the rest saved. The proportion of the additional total income is called the marginal propensity to consume (MPC).

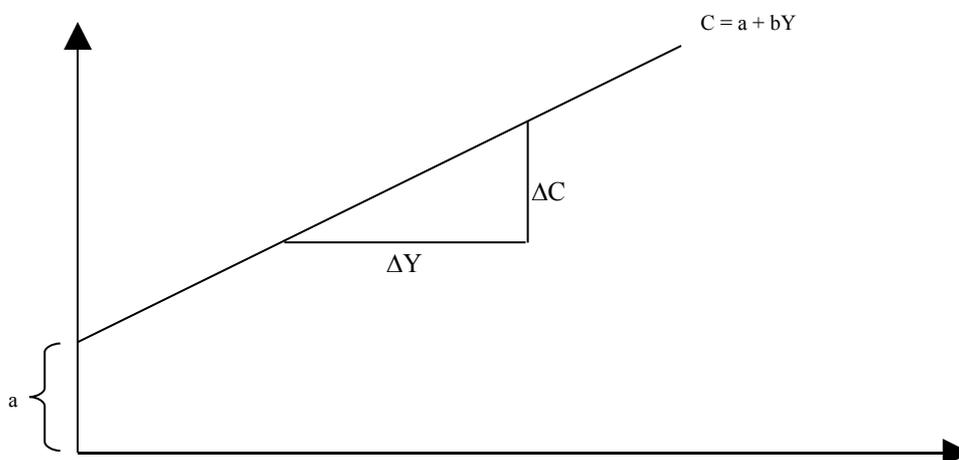
Since in our analysis (ignoring G, T, X and M) saving and consumption are the only two uses for income, $MPC + MPS = 1$.

It is often assumed that the marginal propensity to consume and the marginal propensity to save are constant proportions and that the household will spend (consume):

- (a) a fixed amount of money every period (£a)
- (b) plus a constant percentage of its income (b% of Y)

Similarly a national economy as a whole will spend a fixed amount a plus a constant percentage ($b\%$) of national income Y .

We can then state a consumption function as $C = a + bY$. This can be illustrated diagrammatically (Figure 3 below, in which ΔC represents the increase in consumption arising from an increase of ΔY in the national income)



Give a consumption function $C = a + bY$:

- (a) the marginal propensity to consume is b , where b is the proportion spent on consumption of each extra £1 earned.
- (b) the average propensity to consume will be the ratio of consumption to income, i.e.:

$$\frac{C}{Y} = \frac{a + bY}{Y}$$

For example if an individual household has fixed spending of £100 per month plus extra spending equal to 80% of its monthly income:

- (a) when its monthly income is £800, its consumption will be:
£100 + 80% of £800 = £740
- (b) when its monthly income is £1,000 its consumption will be:
£100 + 80% of £1,000 = £900.

Changes in the marginal propensity to consume and the marginal propensity to save will involve a change of preference by households between current consumption and saving for future benefits. A cause of such a change might be a change in interest rates, which makes the investment of savings more or less attractive than before.

Investment (I)

The total volume of desired investment in the economy depends on such factors as the rate of interest on capital, the marginal efficiency of capital invested, expectations about the future, business confidence and the strength of consumer demand for goods.

The multiplier

Keynes wanted to suggest why the volume of national income might change, by how much, and what would be the consequences of such a change. When national income grows, we have economic growth. Since economic growth might be an economic objective of government, the reasons for economic growth will obviously be of crucial importance for the government's economic planners.

The level of national income might increase or decrease for a number of reasons; for example, there might be a pay rise for workers or an increase in the country's exports. Keynes showed that if there is an initial change in expenditure, say an initial increase in exports, or government spending or investment or consumer spending, a new equilibrium national income level will be reached.

The eventual total increase in national income will be greater in size than the initial increase in expenditure.

This is an important point. A small initial increase in expenditure will result in a bigger total increase in national income before equilibrium is re-established.

The ratio of the total increase in national income to the initial increase in national income is called the multiplier .

$$\text{Multiplier} = \frac{\text{Total increase in national income}}{\text{Initial increase in national income}}$$

The multiplier can thus be defined as a measure of the effect on total national income of a unit change in some component of aggregate demand, in particular, I, G or X i.e. investment spending, government spending or exports.

Numerical illustration of the multiplier

A numerical illustration of the multiplier might help to explain it more clearly. In this example, we shall again ignore taxes, government spending, exports and imports, and assume a simple closed economy in

which all income is either spent on consumption (C) or saved (S). Let's suppose that in this closed economy, marginal propensity to consume (MPC) is 90% or 0.9 i.e. out of any addition to household income, 90% is consumed and 10% saved.

- (a) If income goes up by £200, £180 would be spent on consumption, and £20 saved.
- (b) The £180 spent on consumption increases the income of other people, who spend 90% (£162) and save £18.
- (c) The £162 spent on consumption in turn becomes additional income to others, so that a snowball effect on consumption (and income and output) occurs, as follows:

Stage		<i>Increase in expenditure</i> £	<i>Increase in savings (withdrawals)</i> £
1	Income rises	200.00	-
2	90% is consumed	180.00	20.00
3	A further 90% is consumed	162.00	18.00
4	"	145.80	16.20
5	"	131.22	14.58
	Total increase in income	<u>2,000.00</u>	<u>200.00</u>

In this example, an initial increase in income of £200 results in a final increase in national income of £2,000. The multiplier is 10.

This multiplier is the reciprocal of the marginal propensity to save. Since MPC = 0.9, MPS = 0.1.

$$\text{Multiplier} = \frac{1}{\text{MPS}}$$

$$\text{Increase in national income} = \frac{\text{Initial increase in expenditure}}{\text{MPS}} = \frac{£200}{0.1} = £2,000$$

Note that at the new equilibrium, savings of £200 equal the initial increase in expenditure of £200 but national income has risen £2,000.

If the marginal propensity to consume were 80%, the marginal propensity to save would be 20% and the multiplier would only be 5. Because people save more of their extra income, the total increase in national income through extra consumption will be less.

The multiplier in a national economy works in the same way. An initial increase in expenditure will have a snowball effect, leading to further and further expenditures in the economy.

A 'downward multiplier' or 'de-multiplier' effect also exists. A reduction in investment will have repercussions throughout the economy, so that a small disinvestment (reduction in expenditure/output) will result in a multiplied reduction in national income.

The importance of the multiplier

The importance of the multiplier is that an increase in one of the components of aggregate demand will increase national income by more than the initial increase itself. Therefore if the government takes any action to increase expenditure (eg by raising government current expenditure, or lowering interest rates to raise investment) it will set off a general expansionary process, and the eventual rise in national income will exceed the initial increase in aggregate demand.

This can have important implications for a government when it is planning for growth in national income. By an initial increase in expenditure, a government can 'engineer' an even greater increase in national income, (provided that the country's industries can increase their output capacity), depending on the size of the multiplier.

The size of the multiplier depends on:

- (a) the marginal propensity to save (MPS);
- (b) the marginal propensity to import -because imports reduce national income, and if households spend much of their extra income on imports, the 'snowball' increase in total national income will be restricted because imports are a 'withdrawal' out of the circular flow of income. One of the reasons for a low multiplier in the UK is the high marginal propensity to import that exists in the UK; and
- (c) tax rates -because taxes reduce the ability of people to consume and so are likely to affect the marginal propensity to consume and the marginal propensity to save.

www.allonlinefree.com

Unemployment And Inflation

Your objectives

After completing this chapter you should:

- (a) be aware of some of the main categories of unemployment and their causes
- (b) understand the relationship between unemployment and inflation
- (c) be aware of the main explanations of why inflation occurs
- (d) be aware of the arguments for and against prices and incomes policies

Unemployment

One of the problems that a national economy might face is a high level of unemployment. High unemployment means that there is a large amount of wasted labour resource, and governments will often try to reduce unemployment to an 'acceptable' level.

To understand why unemployment arises, and in order to suggest remedies, it is important to distinguish between a number of different categories of unemployment.

- (a) **Frictional unemployment.** It is inevitable that some unemployment is caused not so much because there are not enough jobs to go round, but because of the friction in the labour market, i.e. the difficulty in quickly matching workers with jobs, caused perhaps by a lack of knowledge about job opportunities. Frictional unemployment occurs where there is a shortage of a given type of worker in one region, but a surplus of the same type in another (eg clerical staff may be plentiful in Wales but in short supply in London). In general, it takes time to match prospective employees with employers, and individuals will be unemployed during the search period for a new job. Frictional unemployment is temporary, lasting for the period of transition from one job to the next.
- (b) **Seasonal unemployment.** This occurs in certain industries, eg building, tourism and farming, where the demand for labour fluctuates in seasonal patterns throughout the year.
- (c) **Structural unemployment** occurs where long-term changes in the conditions of an industry occur, e.g. an industry may decline leaving many workers redundant and reluctant to move to a new industry (labour immobility). The feature of structural unemployment is high regional unemployment in the location of the industry affected.
- (d) **Technological unemployment.** This is a form of structural unemployment, which occurs when new technologies are introduced.

With automation, employment levels in an industry can fall sharply, even when the industry's total output is increasing.

- (e) **Cyclical unemployment.** It has been the experience of the past that domestic and foreign trade go through cycles of boom, decline, recession, recovery, then boom again, and so on.

Cyclical unemployment can be long-term, and a government might try to reduce it by doing what it can to minimise a recession or to encourage faster economic growth.

Seasonal employment and frictional unemployment will be short-term. Structural unemployment, technological unemployment, and cyclical unemployment are all longer-term, and more serious.

Government policies to influence employment

Governments are anxious to influence unemployment levels for a number of reasons. Unemployment is a waste of economic resources, and it leads to individual hardship. Politically, the level of unemployment is seen as a key indicator of a government's success or failure.

A government's policies to influence employment will probably be aimed either at reducing the total number of unemployed people down to an 'acceptable' level, or at creating more jobs. Job creation and reducing unemployment should often mean the same thing, but it is possible:

- (a) to create more jobs without reducing unemployment. This can happen when there is a greater number of people entering the jobs market (school-leavers etc) than there are new jobs being created. For example, if 500,000 new jobs are created during the course of one year, but 750,000 extra school-leavers are looking for jobs, there will be an increase in unemployment of 250,000 even though 500,000 new jobs have been created;
- (b) it is also possible to reduce the official unemployment figures without creating jobs. For example, in the UK, individuals who enrol for one of the government financed Job Training Schemes is taken off the unemployment register, even though he or she does not have a full-time job.

A government can try to create jobs or reduce unemployment by:

- (a) spending more money directly on jobs, i.e. hiring more civil servants;
- (b) encouraging growth in the 'private sector' of the economy. When aggregate demand is growing, firms will probably want to increase output to meet demand, and so will hire more labour;
- (c) encouraging training on job skills. There might be a high level of unemployment amongst unskilled workers, and at the same time a shortage of skilled workers. A government can help to finance training schemes, in order to provide a 'pool' of workers who have the skills that firms need and will pay for.

The Phillips curve

Unemployment and inflation

The problems of unemployment and inflation have been very severe for many countries over recent years.

It has been found from experience that:

- (a) full employment is unachievable without some price inflation, and to increase the level of employment might cause a higher rate of inflation;
- (b) growth in unemployment might sometimes be associated with a rising rate of inflation. Inflationary gaps and deflationary gaps

Inflationary gaps and deflationary gaps

As we saw in Chapter 12 on the Keynesian model, one method of showing equilibrium national income was by means of:

- (a) an aggregate demand curve (AD), where AD is the total demand for all goods in the economy, at different price levels; and
- (b) an aggregate supply curve (AS) where AS is the total supply of all goods in the economy, at different price levels.

The ideas of a deflationary gap and an inflationary gap were also described in Chapter 12. According to Keynes, demand management by the government could be based on government spending:

- (a) to eliminate a deflationary gap and create full employment;
- (b) to eliminate an inflationary gap, to take inflation out of the economy.

A significant point to note from the Keynesian analysis is that when there is unemployment (a deflationary gap) Keynesians accept that reductions in unemployment can only be achieved if prices are allowed to rise, i.e. reducing unemployment goes hand in hand with allowing some inflation.

Relationship between unemployment and inflation: the Phillips curve

A W Phillips discovered (1958) a statistical relationship between unemployment and the rate of money wage inflation which implied that, in general, the rate of inflation fell as unemployment rose and vice versa. A curve, known as a Phillips curve, can be drawn linking inflation and unemployment (Figure 1).

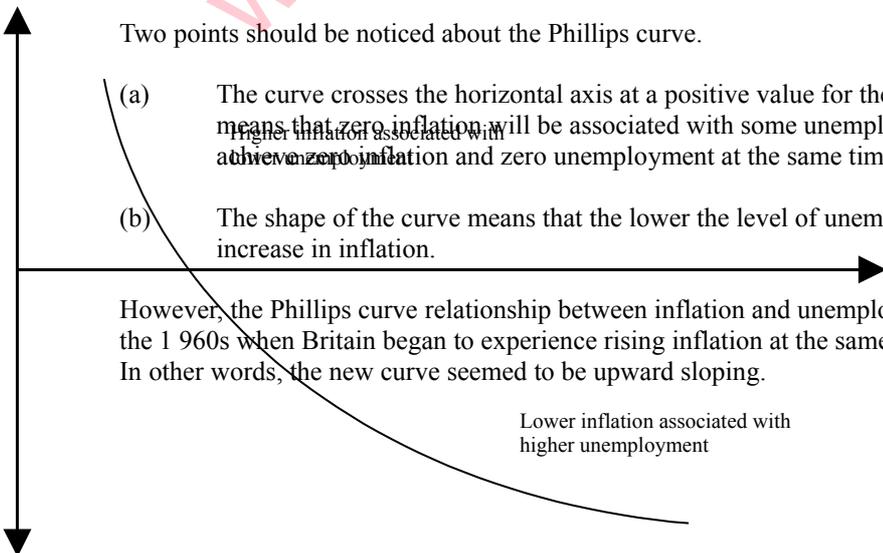
+
0
-

Two points should be noticed about the Phillips curve.

- (a) The curve crosses the horizontal axis at a positive value for the unemployment rate. This means that zero inflation will be associated with some unemployment; it is not possible to achieve zero inflation and zero unemployment at the same time.
- (b) The shape of the curve means that the lower the level of unemployment, the higher the rate of increase in inflation.

However, the Phillips curve relationship between inflation and unemployment broke down at the end of the 1960s when Britain began to experience rising inflation at the same time as rising unemployment. In other words, the new curve seemed to be upward sloping.

Lower inflation associated with higher unemployment



Inflationary expectations: refinements to the Phillips curve

An explanation of rising inflation rates combined with rising unemployment was put forward, based on inflationary expectations. This 'natural rate hypothesis' is supported by monetarist economists.

'Inflationary expectations' are the rates of inflation that are expected in the future. The inflationary expectations of the work force will be reflected in the level of wage rises that is demanded in the annual round of pay negotiations between employers and workers.

- (a) If the work force expects inflation next year to be 3%, they will demand a 3% wage increase in order to maintain the real value of their wages.
- (b) If we now accept that any increase in wages will result in price inflation (which is the monetarist argument), then a 3% pay rise to cover expected inflation will result in an actual rate of inflation of 3%.
- (c) The work force might also try to achieve some increase in the real value of wages. If inflation next year is expected to be 3%, the work force might demand a pay rise of, say, 4%. According to monetarist economists, a pay rise of 4% would simply mean inflation of 4%. If workers wish to achieve a 1% increase in real wages each year, then during each successive period the rate of inflation will begin to accelerate from 5% to 6% to 7% and so on, and the real increases in wages will not happen.
- (d) To compound the problem of inflation still further, it is argued that if mistakes are made over expectations, then money wages will be adjusted upwards next period in order to rectify the mistake made last period.

The example used here is simplified because expectations change and adapt and there are both economic and statutory limits as to changes in both prices and wages. However, many economists believe that events in the UK from about 1967 well into the 1970s followed a sequence of events not unlike that described above.

The natural rate hypothesis

The natural rate hypothesis incorporates these views on inflationary expectations, to produce a refinement of the Phillips curve.

Suppose that the economy is characterised by the Phillips curve PC₁ in Figure 2. Initially, say, that is at an unemployment rate of 4% and zero price and wage inflation.

- (a) Suppose now that the government expands aggregate demand so as to reduce unemployment to, say, 2% of the labour force. There is a movement along the Phillips curve, and the new unemployment level turns out to be associated with 4% inflation.
- (b) As employers realise that they are paying higher wages as well as receiving higher prices, and as workers realise that the real value of their wages has not risen, the unemployment rate rises to 4% again.
- (c) But in the meantime the period of positive inflation has generated inflationary expectations and 4% unemployment is now associated with 3% inflation, because the Phillips curve has shifted from PC₁ to PC₂.

In effect, the short-run Phillips curve has shifted outwards from PC₁ to PC₂ in Figure 2,

Monetarist economists state that the long-run Phillips curve is vertical at the natural rate of unemployment. In our example, monetarists would claim that the long-run Phillips curve is NN in Figure 2 so that there is a natural unemployment rate of 4% (but note that this figure of 4% is simply being used as an example here).

In the long-run, unemployment will revert towards its natural level. The rate of inflation, however, will be determined by the short-run Phillips curve, which will shift upwards as inflationary expectations increase. The distinction between short and long-run Phillips curves can help explain the observation that in Britain unemployment and inflation have often both risen at the same time.

Monetarist economists argue that:

- (a) the only way to reduce the rate of inflation is to get inflationary expectations out of the system. In doing so, excessive demands for wage rises should be resisted by employers;
- (b) a firm approach to reducing the rate of inflation could mean having to accept high levels of unemployment for a while;
- (c) attempts to get the unemployment level below its natural rate (whatever this is) will only result in the long run in higher inflation.

Inflation and its consequences

Inflation has a number of undesirable consequences

- (a) Redistribution of income and wealth. Inflation leads to a redistribution of income and wealth in ways which may be undesirable. Redistribution of wealth might take place from creditors to debtors. This is because debts lose 'real' value with inflation. For example, if you owed £1,000, and prices then doubled, you would still owe £1,000, but the real value of your debt would have been halved.

In general, in times of inflation those with economic power tend to gain at the expense of the weak, particularly those on fixed incomes.

- (b) Balance of payments effects. If a country has a higher rate of inflation than its major trading partners, its exports will become relatively expensive and imports relatively cheap. As a result, the balance of trade will suffer, affecting employment in exporting industries and in industries producing import-substitutes.
- (c) Uncertainty of the value of money and prices. If the rate of inflation is imperfectly anticipated, no one has certain knowledge of the true rate of inflation. As a result, no one has certain knowledge of the value of money or of the real meaning of prices.
- (d) Resource costs of changing prices. A fourth reason to aim for stable prices is the resource cost of frequently changing prices. In times of high inflation substantial labour time is spent on planning and implementing price changes. Customers may also have to spend more time making price comparisons if they seek to buy from the lowest cost source. The term 'menu costs' is sometimes applied to the costs of revising prices; the term alludes to the costs of changing prices on restaurant menus.
- (e) If price rises include rises in the rate of interest, people will hold less of their wealth in the form of cash, and more in the form of interest-bearing assets. Cash holdings will need to be replenished more regularly, and this too causes a financial cost (sometimes referred to as a 'shoe-leather cost', the image being that of an individual wearing out his shoes with frequent trips to the bank).

The causes of inflation

The causes of inflation are complex, because there will be several factors operating simultaneously, each having some effect on price levels.

The causes of inflation might be:

- (a) demand-pull factors;

- (b) cost-push factors;
- (c) expectations;
- (d) growth in the money supply.

Demand-pull inflation occurs when the economy is buoyant and there is a high aggregate demand which is in excess of the economy's ability to supply.

- (a) Because aggregate demand exceeds supply, prices rise.
- (b) Since supply needs to be raised to meet the higher demand, there will be an increase in demand for factors of production, and so factor rewards (wages, interest rates, and so on) will also rise.
- (c) Since aggregate demand exceeds the output capability of the economy, it should follow that demand-pull inflation can only exist when unemployment is low. A feature of inflation in the UK in the 1970s and early 1980s, however, was high inflation coupled with high unemployment.

Traditionally Keynesian economists saw inflation as being caused by demand-pull factors. However, they now accept that cost-push factors are involved as well.

Cost-push inflation occurs where the costs of factors of production rise regardless of whether or not they are in short supply. This appears to be particularly the case with wages: workers anticipate inflation rates and demand wage increases to compensate, thus initiating a wage-price spiral. Interest rate rises can also add to the rate of inflation, because mortgage costs will rise.

A further problem is that once the rate of inflation has begun to increase, a serious danger of 'expectational inflation' will occur. This was discussed earlier in this chapter.

The monetarists argue that inflation is caused by increases in the supply of money. There is a considerable debate as to whether increases in the money supply are a cause of inflation or a symptom of inflation. The monetarists argue that since inflation is caused by an increase in the money supply, inflation can be brought under control by reducing the rate of growth of the money supply.

High interest rates and inflation

A government may adopt a policy of high interest rates as a means of trying to reduce the rate of inflation, when inflation is being caused by a boom in consumer demand (i.e. with demand rising faster than the ability of industry to increase its output to meet the demand).

If interest rates are high enough, there should eventually be a reduction in the rate of growth in consumer spending. This reduction should occur because:

- (a) people who borrow must pay more in interest out of their income. This will leave them less income, after paying the interest, to spend on other things;
- (b) high interest rates might deter people from borrowing, and so there would be less spending with borrowed funds;
- (c) high interest rates should encourage more saving, with individuals therefore spending less of their income on consumption;
- (d) high interest rates will tend to depress the values of non-monetary assets, such as houses, and the reduction in people's perceived wealth may make people feel 'poorer' and consequently reduce the amounts they spend on consumer goods.

Prices and incomes policy

The first prices and incomes policy was introduced in Britain by the then prime minister, Harold Macmillan, in 1962. There have been several attempts to find a successful policy since then, the last being the 'social contract' under the Labour government of James Callaghan.

The need for co-ordinated prices and incomes controls

If prices but not wages are controlled the results would be a profit squeeze, and/or a rise in unemployment, with firms shedding labour in order to remain profitable.

This suggests that wages ought to be kept under control as well as prices. This would be particularly important if wage rises are a major cause of cost-push inflation.

The structure of a prices and incomes policy can take various forms, but it is usually possible to make a clear distinction between types of incomes control and types of price control.

Incomes/wages controls

The broad types of wage controls which could be attempted by a government are:

- (a) statutory restrictions on the amount of wage increases allowed;
- (b) government wage policy guidelines, which it is hoped that wage negotiators will adhere to voluntarily;
- (c) a 'social contract' between the government and trade unions whereby the unions accept a voluntary restraint on wage increases in return for other 'favours' from the government.

Statutory controls are more direct, and if enforceable, more likely to work effectively. Voluntary controls depend on the willingness of trade unions and individual workers to accept the controls, and a social contract, if agreed, is likely to be more morally binding on them than wage policy guidelines.

Problems with wage controls

Problems with wage controls include:

- (a) non-comparable wages between work groups. Some groups are likely to feel 'cheated';
- (b) erosion of pay differentials;
- (c) rewarding productivity improvement\$. Improved productivity should be encouraged, but wage controls might prevent this;
- (d) evading the controls;
- (e) trade union resistance.

Price controls

A policy of controlling wages cannot work unless prices are controlled as well, because workers will not readily accept enforced controls on their earnings when prices are rising more rapidly. In Britain it has always proved to be much easier to control prices than wages. For example, the Price Commission in Britain, until disbanded by the Thatcher government in 1980, had some success in reducing and/or delaying price rises.

A policy for controlling prices also has particular problems to overcome

- (a) The government must decide what prices should be controlled. It is administratively impossible to control all prices charged by all firms, especially small firms, and controls must therefore be applied to selected products, companies or markets, in the hope that if these prices are controlled, other prices will remain in check too.
- (b) It might be necessary to allow exceptions to the rule, so that certain prices can rise in excess of the government's norms or limits.
- (c) Many raw materials, products and services are imported. If the prices of these imported goods go up (e.g. because of a fall in the value of the domestic currency) firms must be allowed to pass on their increased costs into their prices. Rules for permissible price increases must therefore be established.
- (d) If price controls are too rigid, the government might face the problem that a black market economy might develop in certain goods, with prices on the black market rising to levels that are far in excess of the government's limits.
- (e) There is the enormous practical problem of establishing a policing agency or agencies to monitor, control and approve increases in prices (and this applies equally to controls on incomes).

The Demand For And Supply of Money

Your objectives

After completing this chapter you should:

- (a) understand the main features of the monetarist view of money supply and demand
- (b) understand the main features of the Keynesian view of money supply and demand
- (c) understand the implications of these views for interest rates and inflation

The quantity theory of money

There are two broad schools of thought about monetary theory which you need to know about, and economists are not in unanimous agreement. The two broadly differing views are:

- (a) the monetarist view. The monetarist view of money supply and demand, and of the influence of money on interest rates and inflation, derives from the so-called quantity theory of money;
- (b) the Keynesian view. Keynes developed a theory of the demand for money in the 1920s, known as liquidity preference theory, which challenged the quantity theory of money.

The quantity theory of money

Monetarist economists stress the significance of the role of money in the workings of the economy, and base their arguments on the old quantity theory of money.

The classical quantity theory of money was developed by Irving Fisher in 1911 at a time when Britain was on the gold standard, and the quantity of money therefore changed little. It was the generally accepted view until the 1930s about the relationship between the amount of money in the economy and the level of prices. It is a theory about how much money supply is needed to enable the economy to function.

The quantity theory took the view that money was used only as a medium of exchange, to settle transactions involving the demand and supply for goods and services.

This is quite important to remember. Whereas the quantity theory states that the demand for money is simply for spending on foreseeable transactions (and monetarist economists agree with this view) Keynesian economists argue that there are other reasons for wanting to hold money, as we shall see later.

If the number of transactions in the economy is fixed, and independent of the amount of the money supply, then:

The total money value of transactions will be PT

Where P is the price level of goods and services bought and sold; and
 T is the number or quantity of transactions.

The amount of money needed to pay for these transactions will depend on the velocity of circulation. Money changes hands. A person receiving money can use it to make his own purchases. For example, if A pays B £2 for transaction X, B can use the £2 to pay C for transaction Y and C can use the same £2

to pay D for transaction Z. If the three transactions X, Y, and Z all occur within a given period of time then the money value of the transactions is

$$PT = £2, \times 3 \text{ transactions} = £6.$$

The total amount of money is the same £2 in circulation for all three transactions but this money has exchanged hands three times. The velocity of circulation is 3 and $MY = 6$

Where M is the money supply; and
 V is the velocity of circulation.

This brings us to the 'identity' of the quantity theory of money: $MY = PT$.

MV must equal PT because they are two different ways of measuring the same transactions. In practice, the velocity of circulation V is calculated as the balancing figure where:

$$V = \frac{PT}{M}$$

Conclusion from the quantity theory of money

The identity (or 'equation') of the quantity theory of money does not really say very much. It is assumed that M is both the quantity of demand for money and also the money supply.

- (a) An increase in M would reduce V or increase either p or T .
- (b) An increase in V would reduce M or increase either p or T .
- (c) An increase in p would reduce T or increase either M or V .
- (d) An increase in T would reduce p or increase either M or V .

However, three further assumptions can be made.

- (a) V has a roughly constant value. The velocity of circulation of money remains the same at all times, or at least only changes very slowly over time. There were reasons for making this assumption which we need not delve into here. Whether V is constant or not is discussed later
- (b) T is either given or it is independent of the amount of money, M . The reason why T should be a given total was that the supporters of the Quantity theory argued that in the economy, full employment of resources is the norm and if all resources are fully utilised the volume of transactions T must be a constant value.
- (c) The amount of M is determined by other factors and is independent of V , T or (most significantly) P . The money supply could be controlled by government authorities (eg the central bank).

Given these assumptions, the quantity theory of money becomes a theory of price levels because, since $MY = PT$

$$\text{then } P = \frac{MY}{T}$$

If V and T are roughly constant values, p will vary directly with increases or decreases in the amount of M and it is changes in the money supply M that cause prices p to change, not changes in price that cause changes in the money supply. In other words, inflation would be directly related to the money supply, and a 10% increase in the money supply, say, would result in 10% inflation.

The money supply and inflation

There is an argument that the link between the money supply and inflation is a misleading one. But before we look at this counter-argument, let's summarise the points made so far.

Important conclusions from the quantity theory of money equation are that:

- (a) if the velocity of circulation of money, V , is more or less constant, then any growth in the money supply, M , over and above the potential in the economy to increase T , will cause inflation;
- (b) if output in the economy, T , is growing and if the velocity of circulation, V , is constant, then a matching growth in the money supply, M , would be needed to avoid deflation (ie falling prices).

A further conclusion from the quantity theory is that government's monetary policy should be to allow some growth in the money supply if the economy is growing, but not to let the growth in the money supply get out of hand.

The extent to which these conclusions are valid depends largely on whether the velocity of circulation of money is roughly constant or not. For example, if the money supply increases by 10%, and the 'real growth' in the economy -i.e. the increase in the volume of transactions -is 3%, we could predict that inflation will be about 7% but only if the velocity of circulation is constant

Is the velocity of circulation constant?

The velocity of circulation can be measured. It is measured in the UK as the ratio of Gross National Product at current market prices (this is PT) to the average money stock for the quarter (this is M).

The money stock is measured by the various monetary aggregates, eg in the UK, as $M0$, $M2$ and $M4$. There is a different measure of the velocity of circulation for each definition of the money stock.

Measurements of the velocity of the circulation of money in the UK show that V is subject to continual variations up and down. The following table gives some indication of these variations over a six year period.

Velocity of circulation ratios

*Fourth quarter
of year*

Money stock

M0 M4

1983

24.2 1.82

1985

25.8 1.64

1987

28.1 1.47

1989

29.7 1.27

Monetarist economists have argued that in spite of short-term fluctuations, the velocity of circulation V is constant in the long term. Keynesians economists disagree, and argue that V is variable and so there is no direct connection between the money supply and inflation. In the UK, statistics show that the velocity of circulation increased during the 1970s but slowed down during the 1980s except for $M0$.

Even so, it is probably reasonable to suppose that over a longer period of time, V should remain fairly stable.

Another view of the money supply and inflation

It was mentioned earlier that there is a counter-argument to the view that growth in the money supply causes inflation.

The equation or 'identity' $MY = PT$ must be correct, but instead of increases in M causing inflation, some economists would argue that it is inflation which happens first, causing the money supply to rise in step. And if the money supply doesn't increase at once, V will increase instead. The economists who

believe in this argument conclude that the authorities shouldn't expend so much effort on controlling the money supply because they will be tackling a symptom of inflation rather than its cause.

In the same way, if the money supply M is increasing, there might be no change in PT , and indeed, it could be V that is decreasing.

What makes the money supply grow?

We have looked at the demand for and supply of money without yet asking what it is that makes the money supply grow in the first place.

If we define money broadly, to include bank deposits, the main factors that contribute to money supply growth are:

- (a) government borrowing (the Public Sector Borrowing Requirement or PSBR);
- (b) who the government borrows from, banks or non-banks;
- (c) bank lending; and
- (d) flows of money between the country and foreign traders/investors.

The PSBR is usually a positive amount, adding to the growth in the money supply, because in most years, the government has been a net borrower of money. In some periods in the late 1980s however, the UK government borrowed less than the amount of debt it repaid and so there was a negative PSBR. A negative PSBR has a contractionary effect on the money supply.

Approaches to controlling the growth of the broad money supply

The broad approaches that a government might take to control the growth of the money supply, if this were to be part of its monetary policy, are:

- (a) to reduce or control the size of the PSBR;
- (b) to finance as much government borrowing as possible by borrowing from the UK non-bank private sector (e.g. by encouraging National Savings in preference to issuing gilts);
- (c) to control the increase in bank lending. There are various ways in which the government can try to do this, some of which were mentioned in Chapter 9;
- (d) to control external and foreign currency items -e.g. by keeping the balance of payments under control.

The government can pursue a policy of trying to limit the PSBR, perhaps by cutting government expenditures although possibly also by increasing tax revenues. To the extent that the government must borrow, it can try to sell as much debt as possible to the non-bank private sector, and achieve over funding. Both of these policies have been pursued by the UK government in fairly recent times.

The demand for money: what do we mean by 'money'?

We shall now look more closely at the demand for money, which will take us further into the conflicting theories of Keynesian and monetarist economists.

It helps to start by getting a clear idea of what we mean by money. When we consider the theories of demand for money, we take money to mean a non-interest bearing store of wealth.

- (a) Bank notes and current (bank) accounts ('sight deposits') are money.
- (b) Funds in a bank time deposit account are not money. Nor are building society deposits.

Our definition of money has now therefore switched from a broad money definition (M4) to a narrow money definition.

Keynesian views on the demand for money

Liquidity preference

'Liquidity' means assets in the form of cash or 'near-cash', in particular notes and coin and money in a current bank account. Liquidity preference refers to the preference of people to hold on to their savings as money (i.e. in liquid form) rather than investing it.

Keynes used the concept of liquidity preference, which refers to the demand for money, to explain:

- (a) how savings and investment might be temporarily different and;
- (b) how interest rate levels in the economy are arrived at.

In addition, Keynes argued that if households did not hold their savings in the form of money, they would invest it to earn interest. The choice was between money and bonds. (This is a view since disputed by monetarist economists.)

Keynes identified three reasons or motives why people hold wealth as money rather than as interest-bearing securities. These were:

- (a) the transactions motive: households need money to pay for their day-to-day purchases. The level of transactions demand for money depends on household incomes;
- (b) the precautionary motive: people choose to keep money on hand or in the bank as a precaution for a 'rainy day' when it might suddenly be needed;
- (c) the speculative motive: some people choose to keep ready money to take advantage of a profitable opportunity to invest in bonds which may arise (or they may sell bonds for money when they fear a fall in bonds' market prices).

The speculative motive for holding money needs explaining a bit further.

- (a) The reason for holding money instead of investing in bonds is that interest rates are expected to go up. If interest rates go up, bond prices will fall.

For example, if the current market price of bonds which pay 5% interest is £100, and interest rates doubled to 10%, the market value of the bonds would fall, perhaps to £50. So if interest rates are expected to go up, any bonds held now will be expected to lose value, and bond holders would make a capital loss. Thus, it makes sense to hold on to money, for investing in bonds later, after interest rates have gone up.

- (b) What causes individuals to have expectations about interest rate changes in the future?

Keynes argued that each individual has some expectation of a normal rate of interest. This perception of a normal interest rate reflects past levels and movements in the interest rate, and expectations of the future rate level, obtained from available market information.

Keynes argued further that people will need money to satisfy the transactions motive and precautionary motive regardless of the level of interest. It is only the speculative motive which alters the demand for money as a result of interest rates.

- (a) If interest rates are high, people will lend more money (e.g. by buying government stocks) hold little cash, and will have low liquidity preference.

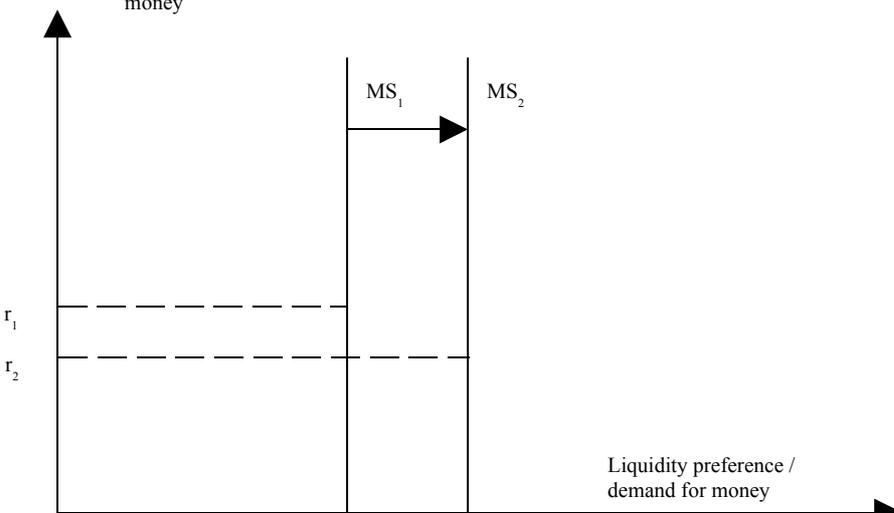
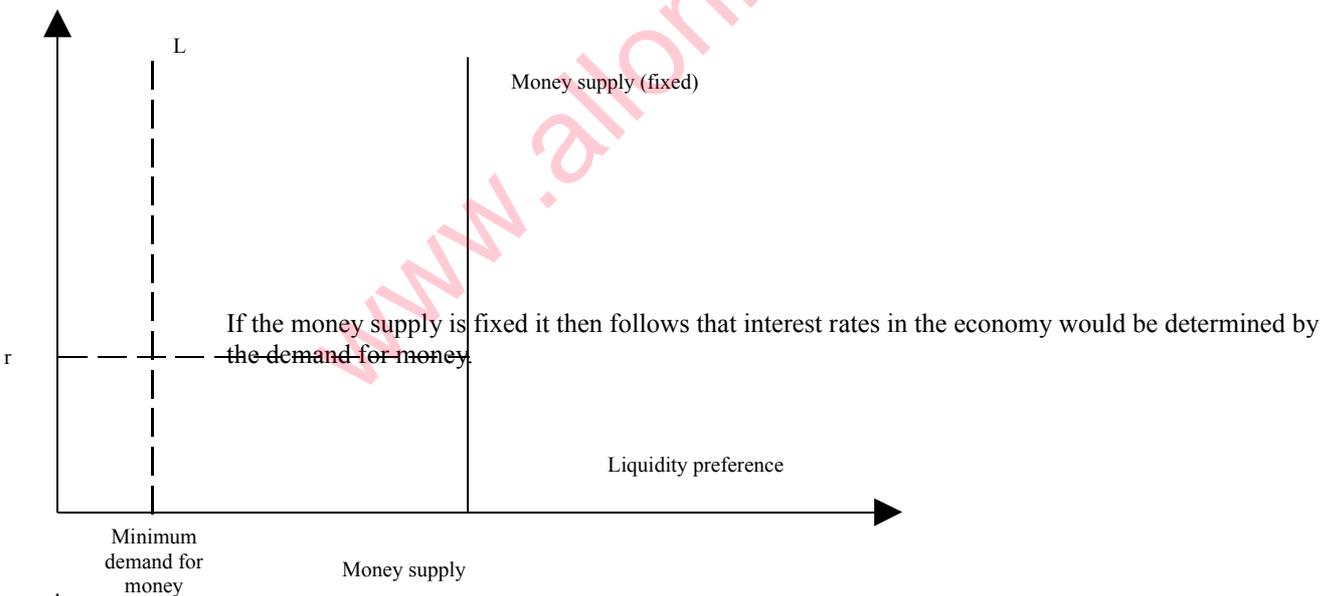
- (b) If interest rates are low and people expect them to rise, there is a danger that current bond prices will fall when interest rates go up. People will therefore hold money to satisfy the speculative motive, i.e. want to invest later in bonds, and their liquidity preference will be high.

The conclusion is that the demand for money will be high (i.e. liquidity preference will be high) when interest rates are low. This is because the speculative demand for money will be high when interest rates are low. Similarly, the demand for money will be low when interest rates are high, because the speculative demand for money will be low.

There is a minimum fixed demand for money (transactions and precautionary motives) and some demand for money that varies with interest rates (speculative motive).

Keynesian view on interest rates and money demand and supply

If the money supply is assumed to be fixed by government decision, the size of the money supply is perfectly inelastic with respect to changes in the rate of interest. Keynes argued that the level of interest rates in the economy would then be reached by the interaction of money supply (fixed) and money demand (liquidity preference), as follows (Figure 1).



If there is an increase in the money supply from M_1 to M_2 in Figure 2, interest rates will fall from r_1 to r_2 . There will be some increase in the level of investment spending, since it now becomes more profitable for firms to invest in new capital.

According to Keynesians, therefore, an increase (or decrease) in the money supply only affects the demand for goods and services (and hence the level of income) indirectly via a change in the rate of interest.

If the rate of interest is at a point where the demand for money is in the so-called liquidity trap, an increase in the money supply would have a very small effect on interest rates.

Investment would then be hardly affected at all. Keynes explained this by saying that the increases in the money supply would be offset by a reduction in the velocity of circulation, so that the increase in the money supply would have a neutral effect on the economy.

Keynesians therefore argued that monetary policy to control the money supply would possibly, though not always, have an effect on interest rates (an increase in the money supply without an increase in the demand for money would make individuals use the extra money available to buy more bonds, and the higher demand for bonds would push down interest rates) and changes in interest rates might in the longer term affect investment. In other respects, however, monetary policy would not really affect the economy and national income, because increases in the money supply would be neutralised by reductions in the velocity of circulation, leaving PT unaffected.

The monetarist viewpoint

Introduction

Monetarist economists such as Professor Friedman revived the quantity theory of money and argued that there is a much more direct link between the money supply and national income.

In other words, whereas Keynes argued that an increase in the money supply would merely result in lower interest rates, with no immediate effect on national income, monetarists argue that an increase in the money supply will lead directly and quickly to changes in national income and PT , with the velocity of circulation V remaining fairly constant.

Friedman's first major work, entitled *The quantity theory of money: a re-statement* was published in 1956.

Friedman argued that money is just one of five broad ways of holding wealth. The five ways are:

- (a) money;
- (b) bonds;
- (c) equities;
- (d) physical goods;
- (e) human wealth.

(Human wealth here is a special concept of wealth and may be ignored for the purpose of our analysis.)

Each method of holding wealth brings some form of return or yield to the holder.

- (a) The yield from money might include some interest, such as on bank deposit accounts, but the main yield from money is a convenience yield. This is the convenience of having ready

money when it is needed instead of having to go to the bother of converting other assets into cash. The convenience yield cannot be measured in money terms.

- (b) Bonds stand for fixed-interest investments. The return on bonds is not just the interest yield but also the capital gain or loss as a consequence of a change in market interest rates.
- (c) Equities are financial assets which should provide a yield (dividends and capital growth) which keeps ahead of the rate of inflation ie a real return to investors.
- (d) Physical assets are houses, consumer durables such as furniture, paintings etc. These are all physical assets which do not waste away through use (because assets which are consumed cannot be a store of wealth).

Friedman argued that the demand for money is related to the demand for holding wealth in its other forms. Money is a direct substitute for wealth in the form of bonds, equities or physical goods. In this respect, he argued against the Keynesian view that holding money is only a substitute for holding financial assets (bonds).

Whereas Keynes believed that if people did not want to hold money, they would invest it to earn interest, monetarists believe that people would possibly invest it to earn interest, but they might also use it instead to buy equities or physical assets.

Friedman argued that money gives a convenience yield but it is not an asset which is held for its own sake. It is a 'temporary abode of purchasing power' waiting to be spent on other types of financial or physical asset. The 'portfolio' of assets held by any person should be a balance between money, bonds, equities and physical assets, i.e. the full range of financial assets and physical assets, so that the marginal interest or yield from the last £1 invested in every type of asset is the same.

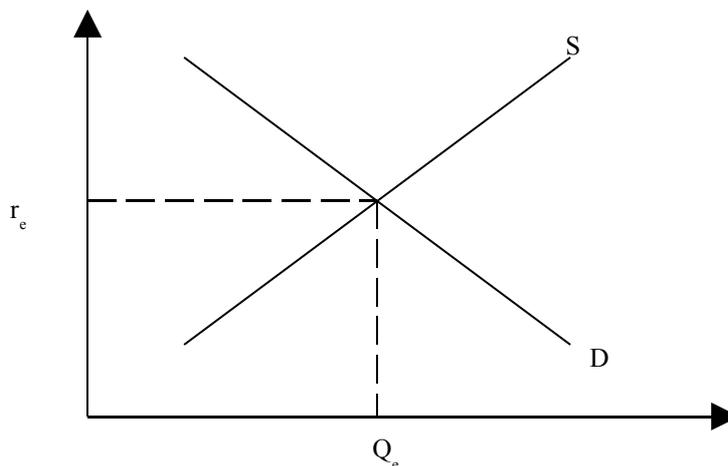
The demand for money is therefore a function of the yield on money and the yield on other forms of holding wealth. Remember, however, that the yield as defined here includes non-monetary yield such as convenience and enjoyment.

Monetarists would argue, further, that the demand for money is fairly interest-elastic. The demand for money is related to a transactions motive, but not to any speculative motive. An expected rise in interest rates might persuade individuals to sell bonds and buy other assets, but not to hold speculative money!

Monetarists and the loanable funds theory of interest rates

According to Keynesians, the level of interest rates is determined by the interaction of the demand for money and the money supply.

Monetarists disagree: after all, they hold the view that the reasons for demanding money are for transactions only, not speculation about future investment. Monetarists argue instead that interest rates are determined by the demand and supply of loanable funds.



An increase in the money supply, without any increase in demand for money, will increase the amount of loanable funds available. Interest rates will fall, and investment will rise.

Monetarists, the transmission and imbalances between money supply and demand

So what is the connection between the demand for money, the money supply and national income, according to monetarists?

This can be explained by the transmission mechanism. The transmission mechanism describes the process whereby any excess of money demand over money supply, or any excess of money supply over money demand, causes a change in the aggregate expenditure in the economy (i.e. a change in national income).

Starting from a position of equilibrium holding of assets of all kinds, an increase in the money supply would leave individuals holding an excess of money balances. In order to restore the level of money holdings to its desired level, individuals will substitute assets of all kinds for money, i.e. the demand for all commodities will increase, not just demand for financial assets.

- (a) The increase in the price of financial assets means a fall in the rate of interest. This will not lead to an increase in the demand for money, according to the monetarists, because they believe that the demand for money is interest-inelastic.
- (b) The increase in direct spending on goods and services will, however, lead to a rise in the level of money national income.

By assuming that money is a substitute for all assets, the monetarists conclude that variations in the money supply have a great influence on the level of national income.

In terms of the quantity theory of money, if M goes up, and $MY = PT$, there will be an increase in PT , but this could mean an increase in either real output (T) or in prices and inflation (P).

Suppose that the demand for money goes up, but the authorities stop the money supply from increasing, so that there is an excess demand for money.

The transmission mechanism will work the other way. Households will sell bonds, equities and reduce consumption on other goods in order to acquire more money. Interest rates will go up. There will be a decline in total spending in the economy until money supply and demand are again in equilibrium. Since $MY = PT$, a decline in PT will have one of two effects:

- (a) if the economy is operating below its full employment national income level, a decline in T (i.e. even less output and so more unemployment);
- (b) if the economy has an inflationary gap, a decline in p (i.e. inflation will be brought under control).

Conclusion. For monetarists, changes in the money supply cause changes in national income. This contrasts with the Keynesian view that changes in the money supply are caused by changes in national income, not vice versa.

The monetarist view of money supply and inflation in the economy

Monetarists argue that since money is a direct substitute for all other assets, an increase in the money supply, given a fairly stable velocity of circulation, will have a direct effect on demand for other assets because there will be more money to spend on those assets. If the total output of the economy is fixed, then an increase in the money supply will lead directly to higher prices.

Monetarists therefore reach the same basic conclusion as the old quantity theory of money. A rise in the money supply will lead to a rise in prices and probably also to a rise in money incomes. (It is also

assumed by monetarists that the velocity of circulation remains fairly constant, again taking a view similar to the old quantity theory.)

In the short run, monetarists argue that an increase in the money supply might cause some increase in real output and so an increase in employment. In the long run, however, all increases in the money supply will be reflected in higher prices unless there is longer-term growth in the economy.

Weaknesses in monetarist theory

There are certain complications with the monetarist views. For example:

- (a) the velocity of circulation is known to fluctuate up and down by small amounts;
- (b) increases in prices will not affect all goods equally. Some goods will rise in price more than others and so the relative price of goods will change. For example, the price of houses might exceed the average rate of inflation but the price of electronic goods might rise more slowly;
- (c) a higher rate of inflation in one country than another might affect the country's balance of payments and currency value, thereby introducing complications for the economy from international trade movements;
- (d) prices in the economy might take some time to adjust to an increase in the money supply.

Nevertheless, monetarists would still argue that inflation has been a major problem for many national economies, including that of the UK. High rates of inflation are impossible without a comparable increase in the money supply. V and T cannot adjust quickly enough, given the identity $MY = PT$. Inflation cannot be brought under control unless the money supply is also controlled.

A comparison of the two views

Which view of money is correct?

Keynesians and monetarists disagree and one key point of difference is the substitutability of money for bonds, with liquidity preference influenced by the rate of interest on bonds. Surely, therefore, it is a simple matter of testing empirically whether this is the case? In other words, all we need to do is to find out whether there is a significant statistical relationship between the quantity of money and the rate of interest. There might be some practical difficulties in setting up the test in the first place, for example defining money and defining bonds, but this should not be beyond the bounds of possibility.

As you might imagine, tests have been carried out to find out whether the direct relationship between money supply and interest rates exists (in which case Keynesians are correct) or does not (in which case monetarists are correct). Unfortunately, empirical tests have not been conclusive. A relationship does exist, but it is not certain whether it is statistically significant or not. Given the view of the monetarists that in the short run interest rates might fall when the money supply is increased, some fall in interest rates after an increase in the money supply would still be consistent with monetarist theory! The Keynesian versus monetarist argument is therefore unresolved.

Both approaches agree that:

- (a) great economic uncertainty explains the high cost of capital in recent years;
- (b) high rates of inflation are largely responsible for high nominal interest rates;
- (c) government monetary policy has some effect on the rate of inflation and so on the cost of capital;
- (d) although they have different views about 'the speed and the mechanism by which interest rate adjustments occur, there is general agreement that interest rates are sometimes rather sticky.

Smooth adjustment of interest rates to changing economic conditions depends on the participants in capital markets recognising the changes that are required'. (Wilson Committee)

It is probably also generally agreed that the interest rate which produces an equilibrium balance between the supply and demand for investment (flow of funds approach) and the interest rate which brings into balance the supply and demand for money (liquidity preference approach) will tend to move towards each other and in the longer term become the same interest rate anyway. The liquidity preference approach, however, is perhaps more concerned with short term movements in interest rates.

www.allonlinefree.com

Fiscal Policy

Your objectives

After completing this chapter you should:

- (a) understand how a government can influence the economy by managing the amounts it spends and the amounts it collects in taxation.
- (b) understand how a government can use taxation to promote its policy aims

Fiscal policy, national income and demand management

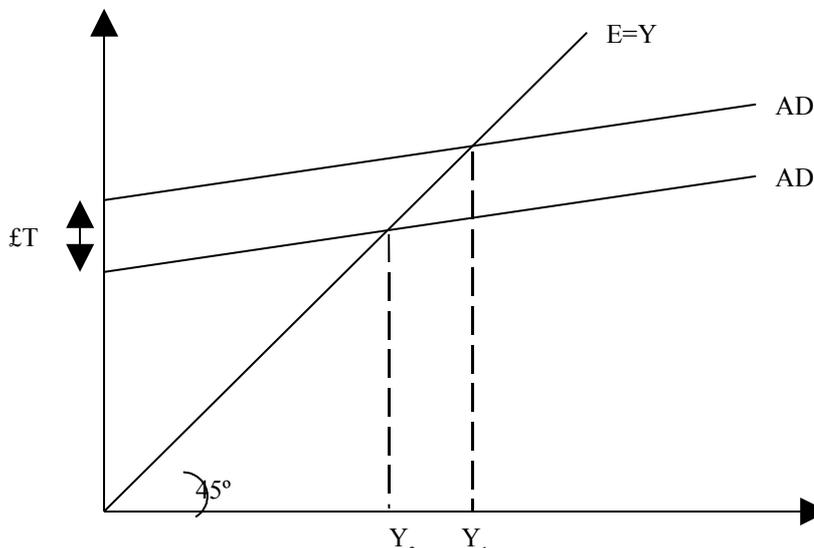
Introduction

The word 'fisc' means the state treasury or the public purse. Fiscal policy relates to matters concerning the state treasury. More specifically, fiscal policy is action by the government to spend money, or to collect money in taxes, with the purpose of influencing the condition of the national economy. Remember that government spending is an injection into the economy, adding to aggregate demand = expenditure = national income, whereas taxes are a withdrawal. A government might intervene in the economy by:

- (a) spending more money and financing this expenditure by borrowing;
- (b) collecting more in taxes without increasing public spending;
- (c) collecting more in taxes in order to increase public spending, thus diverting income from one part of the economy to another.

Demand management, you will remember, is a term used to describe the economic policy of a government when it attempts to influence the economy by changing aggregate demand.

Fiscal policy could be used as an instrument of demand management. Government spending and taxation levels could be used to eliminate an inflationary gap or deflationary gap in the economy. A reduction in taxation would give households a larger income, and so domestic consumption (C) would rise: conversely, an increase in taxation would reduce domestic consumption. Extra government spending (G) should create a multiplier effect on national income, although public sector spending might 'crowd out' private sector investment, because of higher rates of interest in the capital markets.



An increase in taxation by £T, without any matching increase in government expenditure, would reduce the aggregate expenditure in the economy from AD 1 to AD2 and so the money value of national income would fall from y_1 to y_2 . This would result in either a fall in real output or it would dampen inflationary pressures.

Similarly, a reduction in taxation without any reduction in government spending would increase the money value of national income. This would either cause real output to increase, or it would give a boost to price rises and inflation.

Three elements of public finance

Broadly, there are three elements in public finance.

(a) Expenditure

Expenditure by the government, at a national and local level, has several purposes (for example, to fund public services such as a health service, or to pay pensions and unemployment benefits).

(b) Income

Expenditure must be financed, and the government must have income. Most government income comes from taxation. Other income is obtained from direct charges to users of government services (e.g. charges to consumers by nationalised industries, and National Health Service charges).

(c) Borrowing

To the extent that a government's expenditure exceeds its income, it must borrow to make up the difference. The amount that the government must borrow each year is referred to as the Public Sector Borrowing Requirement or PSBR.

Taxation as a deterrent to economic growth

Fiscal policy should be formulated within the guidelines that:

- (a) taxes should be high enough to allow the government to carry on its functions; but
- (b) they should not be set so as to deter private investment and initiative.

Fiscal policy and macroeconomic objectives

Fiscal policy is concerned with government spending and taxation. Refer back to the discussion of the circular flow of income if you find the following analysis unclear.

(a) If government spending is increased, there will be an increase in the amount of injections, expenditure in the economy will rise and so national income will rise (either in real terms, or in terms of price levels only; i.e. the increase in national income might be 'real' or inflationary').

(b) If government taxation is increased, there will be an increase in withdrawals from the economy, and expenditure and national income will fall. A government might deliberately raise taxation to take inflationary pressures out of the economy.

Achieving growth in national income without inflation has been a problem bedeviling governments for many years. Certainly, government spending and government taxation policies can affect economic growth (i.e. the national income level in real terms) but it can also stimulate further inflation.

Fiscal policy can be used to reduce unemployment and provide jobs. For example, more government spending on capital projects would create jobs in the construction industries; lower 'employment taxes'

(such as National Insurance) would possibly make employers more willing to take on extra numbers of employees.

Government spending could however create inflationary pressures, and inflation tends to create more unemployment. Fiscal policy must be used with care, even to create new jobs.

Budget surplus; budget deficit; balanced budget

- (a) It can increase demand directly by spending more itself (e.g. on the health service or education, and by employing more people itself).
 - (i) This extra spending could be financed by higher taxes, but this would reduce spending by the private sector of the economy because the private sector's after-tax income would be lower.
 - (ii) The extra government spending could also be financed by extra government borrowing. Just as individuals can borrow money for spending, so too can a government.
- (b) It can increase demand indirectly by reducing taxation and so allowing firms and individuals more after-tax income to spend (or save).
 - (i) Cuts in taxation can be matched by cuts in government spending, in which case total demand in the economy will not be stimulated significantly, if at all.
 - (ii) Alternatively, tax cuts can be 'financed' by more government borrowing.

Just as aggregate demand in the economy can be boosted by either more government spending or by tax cuts, financed in either case by a higher PSBR, so too can demand in the economy be reduced by cutting government spending or by raising taxes, and using the savings or higher income to cut government borrowing.

Expenditure changes and tax changes are not mutually exclusive options, of course. At the time of writing (August 1993), the UK Government is considering a package of measures, possibly to include tax increases and reduced Government spending, in order to reduce the current high PSBR.

When a government's income exceeds its expenditure, and there is a negative PSBR, i.e. a Public Sector Debt Repayment or 'PSDR' we say that the government is running a budget surplus.

When a government's expenditure exceeds its income, so that it must borrow to make up the difference, there is a PSBR and we say that the government is running a budget deficit.

When a government's expenditure and income are the same, so that the PSBR is nil, there is a balanced budget.

Fiscal policy and the Budget

A government must plan what it wants to spend, and so how much it needs to raise in income or by borrowing. It needs to make a plan in order to establish how much taxation there should be, what form the taxes should take and so which sectors of society (firms or households, the rich or the poor etc) the money should come from. This formal planning of fiscal policy is usually done once a year. With effect from 1993, the annual Budget statement is to take place in November or December, and will include the Government's plans for both spending and taxation.

This annual review of taxation means that a full review of the government's fiscal policy can only be done once a year. In between Budgets, a government must resort to other non-fiscal policy instruments to control the economy, such as influencing interest rate levels.

The National Debt

The National Debt is the amount of debt owed by the central government of a country to its various creditors. Creditors may be nationals of the country (e.g. investors in government loan stock) or foreign nationals (perhaps foreign banks or even the International Monetary Fund).

When the National Debt is high, interest repayments will account for a large proportion of central government expenditure. It is therefore important that when a government borrows money, it should be invested or spent in such a way as to ensure that sufficient income is eventually generated to repay the interest and debt capital. If a government is unable to do this (as in the case of many developing and less developed countries at the moment) it will lose its creditworthiness and find future loans much harder to come by.

The ability to repay is probably the major factor in setting an effective limit to the size of a national debt.

The National Debt exists in the form of debt 'instruments' which consist of:

- (a) marketable debt (mainly long-term debt, which consists of gilt-edged securities, which are sold to investors and traded on The Stock Exchange); and
- (b) non-marketable debt (including National Savings and any non-marketable loans raised by the government).

To service the National Debt, a government must:

- (a) pay interest on the debt; and
- (b) make capital repayments when they fall due

However, the problem of making capital repayments can be overcome, if required, by taking out new loans when old loans mature (e.g. to repay a loan of £ 100 million, a government can borrow a further £100 million).

If the loan is obtained from the private sector of the country's economy, the National Debt and servicing the debt involve the transfer of funds between different sections of society. When the government borrows, it takes money from one group and spends it on other sections of society. When the government pays interest, it will raise money in taxes from the rest of society and pay this money to its debtors. Once again, this involves a transfer of the funds.

The National Debt and servicing the National Debt thus involve:

- (a) a redistribution of funds within society, through government borrowing and spending, or through taxation to pay debt interest;
- (b) borrowing to spend 'now' and only repaying the debt with interest 'later'. In other words, society benefits now, and the payment burden falls on society in later years, perhaps in some cases as long as a generation or so later.

When money is borrowed from abroad, the flow of funds in obtaining the money and in repaying the debt crosses national boundaries, and so there are implications for the balance of payments and the exchange rate of the domestic currency.

Public expenditure control and the PSBR

The public sector borrowing requirement or PSBR is the annual excess of spending over income for the entire public sector, not just central government. If the entire public sector had a balanced budget, the PSBR would be nil. When there is an annual excess of income over expenditure, as there was in the UK a few years ago, some of the National Debt can be repaid and there is a Public Sector Debt Repayment (PSDR).

Government spending on both revenue and capital items has to be financed by either:

- (a) taxation and other 'current' revenues; or
- (b) borrowing.

Keynesian economists argue the need for more government spending in a recession to boost demand in the economy and create jobs. An argument put forward is that it makes sound financial sense to use current revenues from taxation etc. to meet current expenditures and interest payments on debt, but to use borrowing for capital expenditure needs. Since taxation revenue exceeds current spending, the argument concludes that the government should borrow more for the capital spending.

The UK government rejected this argument, and believed that the upward spiral of public expenditure has made the burden of taxation on companies and individuals intolerable. 'There can be no prospect of bringing the burden of tax back to tolerable levels without firm control over public expenditure growth, by deciding first what can be afforded then setting plans for individual programmes consistent with that decision.' (The Chancellor's Autumn Statement, November 1983).

The central government in the UK has also been greatly concerned with local authority spending, and its policies to control the growth of spending have raised a major political storm between the government and opposition parties, especially in view of the large number of Labour-controlled councils.

Taxation policy and its effects

Introduction

The functions of taxation include the following:

- (a) to raise revenues for the government, local authorities and similar bodies (e.g. the European Community). The revenues are used to provide goods and services that the market economy either does not provide at all (e.g. defence) or will not provide in sufficient quantities (e.g. education) and to pay for the upkeep of government administration;
- (b) to discourage certain activities regarded as undesirable. The imposition of Development Land Tax in the United Kingdom in the mid- 70s was partially in response to the well-publicised growth in property speculation;
- (c) to cause certain products to be priced to take into account their social costs. For example, smoking entails certain social costs, including the cost of hospital care for those suffering from smoking-related diseases, and so the government sees fit to make the price of tobacco reflect these social costs;
- (d) to redistribute wealth. The higher rates of income tax up to 1979 and capital transfer tax (now replaced by inheritance tax) were designed to transfer wealth from the better off to the less well off through higher social security benefits. Some politicians favour a wealth tax;
- (e) to protect industries from foreign competition. If the government levies a duty on all imported goods much of the duty will be passed on to the consumer in the form of higher prices, making imported goods more expensive. This has the effect of transferring a certain amount of demand from imported goods to domestically produced goods;
- (f) to provide a stabilising effect on national income. Taxation reduces the effect of the multiplier, and so can be used to dampen upswings in a trade cycle -ie higher taxation when the economy shows signs of a boom will slow down the growth of money GNP, and so take some inflationary pressures out of the economy.

Taxation can be used to achieve the same purposes as government expenditure. In addition, it can be used to affect particular sectors of the economy. For example, taxes on wealth, capital and income are potentially effective tools for redistribution of wealth.

Qualities of a good tax

Adam Smith in his *Wealth of Nations* (1776) ascribed four features to a good tax system.

- (a) Persons should pay according to their ability.
- (b) The tax should be 'certain' and easily understood by all concerned.
- (c) The payment of tax should ideally be related to how and when people receive and spend their income (e.g. PAYE is deducted when wages are paid, and VAT is charged when goods are bought).
- (d) The cost of collection should be small relative to the yield (by this criterion, the car road tax is an inefficient tax).

Regressive, proportional and progressive taxation may be distinguished.

- (a) A regressive tax takes a higher proportion of a poor person's salary than of a rich person's. Television licences and road tax are examples of a regressive tax since they are the same for people of all wealth classes. A poll tax or community charge is also a regressive tax because it takes a greater proportion of the income of a poor person than a rich person, even though individual charges can vary from one local authority to another, and in spite of some rebates for the least well-off groups.
- (b) A proportional tax takes the same proportion of income in tax from all levels of income. Schedule E income tax with a basic of tax at 25% is proportional tax, but only within a limited range of income.
- (c) A progressive tax takes a higher proportion of income in tax as income rises. Income tax as a whole is progressive, since the first part of an individual's income is tax free due to personal allowances and the rate of tax increases in steps from 20p in £1 to 40p in £1 as taxable income rises.

Advantages of progressive taxation

There are several arguments in favour of progressive taxes.

- (a) They are levied according to the ability of individuals to pay. Individuals with a higher income are more able to afford to give up more of their income in tax than low income earners, who need a greater proportion of their earnings for the basic necessities of life. If taxes are to be raised according to the ability of people to pay (which is one of the features of a good tax suggested by Adam Smith) then there must be some progressiveness in them.
- (b) Progressive taxes enable a government to redistribute wealth from the rich to the poor in society. It is likely that there will be little redistribution of wealth among the middle-income ranges of society, but there should be some redistribution of wealth from the richest members of society to the poorest. Such a redistribution of wealth might be regarded as a matter of social justice but it will also alter the consumption patterns in society since the poorer members will spend their earnings and social security benefits on different types of goods than if the income had remained in the hands of the richer people.
- (c) Indirect taxes (e.g. taxes on expenditure) tend to be regressive and progressive taxes are needed to counterbalance regressive taxes in the tax system, and so make the tax system as a whole more fair.

The disadvantages of progressive taxation

There are some arguments against progressive taxes.

- (a) Some would argue that in an affluent society there is less need for progressive taxes than in a poorer society, and there is less need to redistribute wealth from the very wealthy to the fairly well-to-do.
- (b) When progressive taxes are harsh, and either tax high income earners at very high rates on marginal income, or tax the wealthy at high rates on their wealth, it is argued by some that they will act as a deterrent to initiative.
- (c) It is also sometimes argued that individuals and firms suffering from high taxes might try:
 - (i) to find loopholes so as to avoid paying tax (e.g. non-taxable perks);
 - (ii) to evade taxes, i.e. withhold information about their income or wealth from the authorities;
 - (iii) to transfer their wealth to other countries, or establish companies in tax havens where corporate tax rates are low.

But of course some people will try to avoid or evade tax whether it is high or not so high.

Advantages and disadvantages of a proportional tax

The advantage of a tax which is proportional to income is primarily that of fairness.

The disadvantages of a proportional tax are"

- (a) the large administrative system needed to calculate personal tax liabilities on a proportional basis. Income tax, for example, can be a costly tax to collect where individuals (especially self-employed people) require detailed tax assessments;
- (b) the tax rules may need to be quite complex in order to be proportional;
- (c) they do not contribute towards a redistribution of wealth among the population.

Advantages and disadvantages of a regressive tax

The main disadvantage of a regressive tax is that it is not equitable, because a greater tax burden falls on those least able to afford it.

An advantage of a regressive tax is that it can be relatively easy to administer and collect. However, a regressive tax could also be expensive to collect. Arguments against a poll tax (community charge) in the UK have been that it is both regressive (and so unfair) and also expensive to administer .

Direct and indirect taxes

A direct tax is paid direct by a person to the Revenue authority. Examples of direct taxes in the UK are income tax, corporation tax, capital gains tax and inheritance tax. A direct tax can be levied on income and profits, or on wealth. Direct taxes tend to be progressive or proportional taxes. They are also usually unavoidable, which means that they must be paid by everyone.

An indirect tax is collected by the Revenue authority (e.g. Customs & Excise) from an intermediary (e.g. a supplier) who then passes on the tax to consumers in the price of goods they sell. Indirect taxes include V A T , excise duty on spirits and beer, and customs duties on goods imported from countries outside the Common Market. Indirect taxes tend to be regressive (unless they are charged exclusively on luxury items). They are also usually avoidable, which means that people can choose not to pay the tax, by not buying the goods or services on which the tax is levied.

Direct taxation: advantages and disadvantages

The main advantages of direct taxes on income are that they are fair and equitable because they are usually progressive or proportional taxes. They can be levied according to ability to pay. Because of their generally progressive nature they also tend to stabilise the economy, automatically taking more money out of the system during a boom and less during depression. Moreover, because they are more difficult to pass on, they are less inflationary than indirect taxes. Taxpayers also know what their tax liability is.

However, taxes on income are often criticised for causing harmful distortions, especially on incentives to work. But as already indicated, empirical research is not conclusive on this matter.

High marginal rates of tax may encourage tax avoidance (i.e. finding legal loopholes in the tax rules so as to avoid paying tax). In Britain, this led to a growth of making income payments-in-kind by way of fringe benefits, for example, free medical and life insurance, preferential loans and favourable pension rights. The government has responded to this by bringing an increasing number of fringe benefits into the tax net (for example, the private use of a company car).

A direct tax on profits is likely to act as a disincentive to risk-taking and enterprise. The tax will reduce the net return from a new investment and any disincentive effects will be greater when the tax is progressive. In addition, a tax on profits will reduce the ability to invest. A considerable part of the finance for new investment comes from retained profits so any tax on corporate profits will reduce the ability of firms to save and therefore limit the sources of funds for investment.

Indirect taxation: advantages and disadvantages

Indirect taxes are 'hidden' in the sense that the taxpayer is largely unaware of the amount of tax he is paying (e.g. the tax on beer and spirits). This has considerable advantages from the government's point of view.

Indirect taxes are generally more difficult to evade than direct taxes

It is sometimes argued that indirect taxes are preferable to direct taxes in that they leave the individual with a choice of not paying the tax by not consuming the taxed commodity.

Indirect taxation can be used to encourage or discourage the production or consumption of particular goods and services and may hence affect the allocation of resources. For example, the production of goods that produce environmental pollution may be taxed as a means of raising the price in order to reduce demand and output. Similarly, the consumption of cigarettes can be discouraged by high indirect taxation. The consumption of imported goods can be discouraged by high import duties, which add to their price in domestic markets.

Indirect taxation is a flexible instrument of economic policy. The rates of indirect taxes may be changed quickly and easily and can take effect immediately. For example, if the Chancellor wished to boost aggregate demand in the economy he could reduce rates of indirect taxation and this would have a rapid effect on private consumption.

Indirect taxes do have disadvantages however

- (a) They can be inflationary. In the UK, a switch from direct tax to indirect tax in the budget of 1979, when V A T rates were increased, resulted in a large increase in the reported rate of inflation. This is, however, largely a one-off effect.
- (b) As already stated, indirect taxes tend to be regressive,
- (c) Indirect taxes are not impartial in their application in other ways. For example, someone who seeks to relax with a cigarette in a pub is going to be much more heavily hit by indirect taxes than someone who likes walking and reading.
- (d) The disincentive effects of indirect taxes are often claimed to be less than those of direct taxes, largely because indirect taxes are not as visible as direct taxes and are therefore less likely to be taken into account in decision-making.

- (e) As we saw in an earlier chapter, firms will maximise their profits at a lower volume of output than if there were no taxes.

The incidence of taxation

The incidence of a tax is the distribution of the tax burden. A tax's formal incidence can be distinguished from its actual incidence.

- (a) In the case of income tax, formal and actual incidence are the same unless workers negotiate a higher post-tax wage, in which case employers bear some of the tax burden. Employers might then raise their product price and pass some of the tax burden on to consumers.
- (b) Corporation tax falls formally on profits and so on shareholders; but the tax could be passed forward to consumers in the form of a price rise following on a reduction in supply and the actual incidence of the tax would then fall on the consumers.
- (c) According to economic theory, producers will not pass the burden of a profit tax on to consumers because they are already producing at a profit-maximising level of output and the marginal condition $MR = MC$ is not affected by a tax. However, it might be more plausible to suppose that producers would regard the tax as an additional cost, in which case it would justify a reduction in output and rise in price. The consumers would in this case bear some of the burden.
- (d) A general indirect tax on all goods would be borne fully by consumers, since they cannot buy alternative goods which are not taxed. VAT is the closest equivalent to a general indirect tax in Britain. Only necessities are exempt from VAT or zero-rated, eg food, fuel and power (for the time being), transport, health services -and books!

Taxation and the redistribution of wealth

Another aspect of taxation is the extent to which it can be used to redistribute wealth, and the consequences this would have for economic activity. Suppose, for example, that a country has no taxation, and most national income is earned by a very small and wealthy proportion of the population. The expenditure pattern within the country would probably be:

- (a) fairly substantial spending by the wealthy few on luxury goods and specialist services (eg domestic servants);
- (b) substantial spending by the poor majority on cheap basic necessities for living, such as simple food and cheap clothing.

Now suppose that a progressive income tax is imposed, which successfully redistributes wealth among the population, so that general income levels rise. The result should be a general rise in living standards, with:

- (a) greater demand for some items; and
- (b) some switch in demand from cheap goods to better quality, more expensive substitutes.

International Trade

Your objectives

After completing this chapter you should'

- (a) understand the theory of comparative advantage and its implications for the development of international trade;
- (b) understand the kinds of restrictions which may be applied to the free movement of international trade, and the reasons for and against adopting such measures.

Absolute and comparative advantage

Reasons for international trade

In the modern economy, production is based on a high degree of specialisation. Within a country individuals specialise, factories specialise and whole regions specialise. Specialisation increases productivity and raises the standard of living. International trade extends the principle of the division of labour and specialisation to countries. International trade originated on the basis of nations exchanging their products for others which they could not produce for themselves. Britain for example imports tea and coffee and exports oil to non-oil producing countries.

International trade arises for a number of reasons,

- (a) Different goods require different proportions of factor inputs in their production,
- (b) Economic resources are unevenly distributed throughout the world.
- (c) The international mobility of resources is extremely limited.

Since it is difficult to move resources between nations, the goods which 'embody' the resources must move. Hence nations which have an abundance of land relative to labour will concentrate on land intensive commodities such as agricultural products. These will be exchanged for 'labour-intensive' products such as manufactured goods made by countries which have an abundance of labour and capital relative to land. The main reason for trade therefore is that there are differences in the relative efficiency with which different countries can produce different goods and services.

Economists distinguish the concepts of comparative advantage and absolute advantage. To explain this distinction we make the following assumptions in what follows.

- (a) There are only two countries, country X and country y
- (b) Only two goods are produced, lorries and wheat.
- (c) There are no transport costs and no barriers to trade.
- (d) Resources within each country are easily transferred from one industry to another.

Absolute advantage

A country is said to have an absolute advantage in the production of a good when it is more efficient than another country in the production of that good, i.e. when it can produce more of a particular good with a given amount of resources than another country. It is a common situation for one country to be more efficient than another in the production of a particular good.

Assuming that y produces wheat more efficiently than country X, while country X has an absolute advantage in producing lorries, a simple arithmetical example can illustrate the potential gains from trade. The table below shows the amounts of lorries and wheat that each country can produce, assuming that each country has an equal quantity of resources and devotes half of its resources to lorry production and half to wheat production.

	<i>Lorries</i>	<i>Wheat (tons)</i>
Country X	20	100
Country Y	<u>10</u>	<u>150</u>
World total	<u>30</u>	<u>250</u>

The relative or comparative cost of lorry production is lower in country X than country Y, but the situation is reversed in the case of wheat production. Country X has an absolute advantage in lorry production and country y has an absolute advantage in wheat production.

Greater specialisation will, however, increase total output of both lorries and wheat.

In order to obtain the benefits of specialisation these countries must exchange some part of their individual outputs. It is not possible to specify the exact rate of exchange but the limits of the exchange rate must be somewhere between the domestic opportunity cost ratios of the two countries. One country will not benefit from international trade if the exchange rate is not between these ratios.

Comparative advantage

When two countries produce the same two goods, as in the example above, and each has an absolute advantage in the production of one good, then it is easy to show that specialisation will lead to an increase in their combined output. Specialisation and trade can still be mutually advantageous, however, even if one country has an absolute advantage in the production of both goods. This will be the case if each country has a comparative advantage in the production of one good. This is summed up in the law of comparative advantage (or comparative costs) which states that two countries can gain from trade when each concentrates on the production of that good in which it has greatest comparative advantage. Comparative cost relates to the opportunity costs of producing the goods and not the absolute costs.

The principle of comparative costs can be shown by an arithmetical example. It is now assumed that country X is more efficient in the production of both lorries and wheat. If each country devotes half its resources to each industry the assumed production totals are shown below.

	<i>Lorries</i>	<i>Wheat (tons)</i>
Country X	20	200
Country Y	<u>10</u>	<u>150</u>
World total	<u>30</u>	<u>350</u>

In terms of resources used, the costs of production in both industries are lower in country X. If we consider the opportunity costs, however, the picture is rather different. In country X the 'cost' of one lorry is ten tons of wheat, i.e. in devoting resources to the production of one lorry in country X there is a sacrifice in terms of ten tons of wheat forgone. The opportunity cost of one lorry in country y is fifteen tons of wheat. Country X therefore has a comparative advantage in the production of lorries. In country X the cost of a ton of wheat is 1/10 of a lorry, while in country y the cost is 1/15 of a lorry. In terms of the output of lorries forgone, wheat is cheaper in country y than in country x. Country y has a comparative advantage in wheat. If each country completely specialises in the production of the good where it has a comparative advantage, the figures below show that total output of lorries increases, but the total output of wheat falls.

	<i>Lorries</i>	<i>Wheat (tons)</i>
Country X	40	0
Country Y	<u>0</u>	<u>300</u>
World total	<u>40</u>	<u>300</u>

The figures above show that the world total of lorry production has risen by ten lorries but that the world production of wheat has fallen by fifty tons. It is possible to show that the increase in the output of lorries, in value terms, more than offsets the fall in the output of wheat. This is not necessary, however, because by only partially specialising in the more efficient country -country X -it is possible to have more of both commodities.

Total output of both goods is greater than that which was obtained when both countries were producing only for domestic consumption. Since the opportunity cost ratios are different in the two countries, beneficial trade is possible. If the opportunity cost ratios were the same in the two countries, the countries would not benefit from specialisation and international trade.

Generalising from this example, we can state the law of comparative advantage:

Each country should export goods in which it has a comparative cost advantage.

In the example above we have assumed that transport costs for international trade are negligible. High transport costs, however, can negate the advantages of specialisation and international trade.

Does the law apply in practice?

The law of comparative advantage does apply in practice, and countries do specialise in the production of certain goods. However, there are certain limitations or restrictions on how it operates.

- (a) Free trade does not always exist. Some countries take action to protect domestic industries and discourage imports. This means that a country might produce goods in which it does not have a comparative advantage.
- (b) Transport costs can be very high in international trade so that it is cheaper to produce goods in the home country rather than to import them.
- (c) Countries might produce similar goods, but give them sufficiently unique characteristics to make them competitive in any country of the world. For example, companies in France, Italy, Spain, Germany and the UK all manufacture cars, but there are sufficient varieties of design that each country can export its own cars to the other countries. For example, the UK imports Fiats, Volkswagens and Citroens, but exports Rover cars to other EC countries.

Free trade and its advantages

The main advantages

The law of comparative advantage states perhaps the major advantage of encouraging international trade. However, there are other advantages to the countries of the world from encouraging international trade. These are as follows.

- (a) Some countries have a surplus of raw materials to their needs, and others have a deficit. A country with a surplus (e.g. of oil) can take advantage of its resources to export them.

A country with a deficit of a raw material must either import it, or accept restrictions on its economic prosperity and standard of living.
- (b) International trade increases competition amongst suppliers in the world's markets.
- (c) Greater competition reduces the likelihood of a market for a good in a country being dominated by a monopolist. The greater competition will force firms to be competitive and so will increase the pressures on them to be efficient, and also perhaps to produce goods of a high quality.

- (d) International trade creates larger markets for a firm's output, and so some firms can benefit from economies of scale by engaging in export activities. Economies of scale improve the efficiency of the use of resources, reduce the output costs and also increase the likelihood of output being sold to the consumer at lower prices than if international trade did not exist.
- (d) There are political advantages to international trade, because the development of trading links provides a foundation for closer political links. An example of the development of political links based on trade is the European Community.

Free movement of capital

Free trade is associated with the free movement of goods (and services) between countries. Another important aspect of international trade is the free movement of capital.

- (a) If a UK company or investor wishes to set up a business in a different country, or to take over a company in another country, how easily can it transfer capital from the UK to the country in question, to pay for the investment?
- (b) Similarly, if a Japanese company wishes to invest in the UK, how easily can it transfer funds out of Japan and into the UK to pay for the investment?

Some countries (including the UK in recent years) have allowed a fairly free flow of capital into and out of the country. Other countries have been more cautious, mainly for one of two reasons.

- (a) The free inflow of foreign capital will make it easier for foreign companies to take over domestic companies. There is often a belief that certain 'key' industries should be owned by residents of the country. Even in the UK, for example, there are restrictions on the total foreign ownership of shares in companies such as British Aerospace and Rolls Royce.
- (b) Less developed countries especially, but other more advanced economies too, are reluctant to allow the free flow of capital out of the country. After all, they need capital to come into the country to develop the domestic economy.

For countries with a large and continuing balance of trade deficit, such as the UK and the USA, it is essential that capital should come into the country to finance the deficit. The balance of payments is discussed in detail in the next chapter.

Free trade and protection

Introduction

Free trade exists where there is no restriction on imports from other countries or exports to other countries. In practice, however, many barriers to free trade exist because governments wish to protect home industries against foreign competition.

Protectionist measures

Protection can be applied by a government in several ways.

They include:

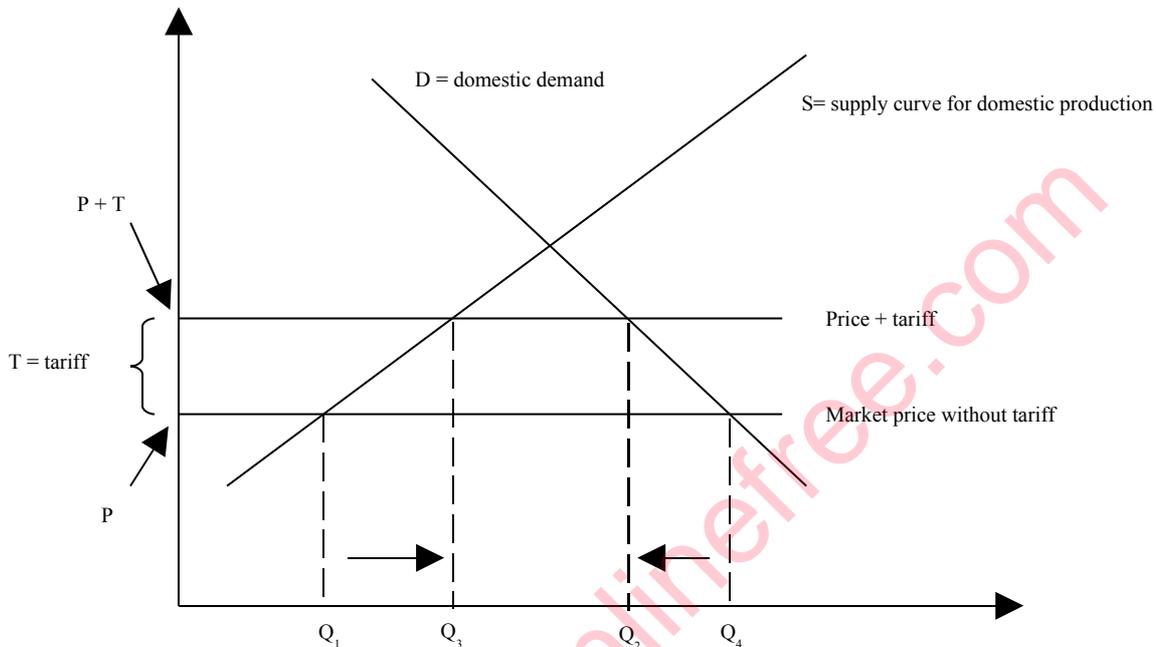
- (a) tariffs or customs duties;
- (b) import quotas;
- (c) embargoes;
- (d) hidden subsidies for exporters and domestic producers;
- (e) import restrictions;
- (f) government action to devalue the nation's currency -ie to reduce its foreign exchange value.

www.allonlinefree.com

Tariffs or customs duties

Tariffs or customs duties are taxes on imported goods. The effect of a tariff is to raise the price paid for the imported goods by domestic consumers, while leaving the price paid to foreign producers the same, or even lower. The difference goes to the government.

For example, if goods imported to the UK are bought for £100 per unit, which is paid to the foreign supplier, and a tariff of £20 is imposed, the full cost to the UK buyer will be £120, with £20 going to the government.



In Figure 1 the market purchase price of the good is P . At that price, domestic suppliers are willing to supply Q_1 but consumers are willing to buy Q_4 . The difference $Q_4 - Q_1$ is then the amount of imports.

An import tariff will raise the price to the consumer to $p + T$. The domestic suppliers need not, of course, raise their prices, but at the higher price, consumers demand Q_2 .

If the domestic producers were to raise their prices to $p + T$ then they would expand their output to Q_3 . Imports would fall from $(Q_4 - Q_1)$ to $(Q_2 - Q_3)$.

The end result of imposing the tariff is that:

- domestic consumers buy fewer units;
- domestic producers supply more to the market;
- foreign suppliers provide less to the market;
- the government earns some tax revenue.

In such cases, import duties benefit the domestic producers and the government, but they harm the consumer. The government raises revenue and domestic producers expand their sales, but consumers either pay higher prices if they buy imported goods, or are forced to buy domestic goods. In the latter case there must be some loss of welfare because previously consumers were not buying domestically-produced goods at the world price, P . Now they are forced to because of the higher prices, $p + T$.

Note that the price elasticity of the demand and supply functions may be important in determining by how much a tariff will reduce imports. If demand and supply are price inelastic, a tariff will have a fairly small effect on import volumes.

Import Quotas

Import quotas are restrictions on the quantity of a product that is allowed to be imported into the country. The quota has a similar effect on consumer welfare to that of import tariffs, but the overall effects are more complicated.

- (a) Both domestic and foreign suppliers enjoy a higher price, while consumers buy less at the higher price.
- (b) Domestic producers supply more
- (c) There are fewer imports (in volume).
- (d) The government collects no revenue.

An embargo on imports from one particular country is a total ban, i.e. effectively a zero quota.

Hidden export subsidies and import restrictions

An enormous range of government subsidies and assistance for exports and deterrents against imports have been practised, such as:

- (a) for exports -export credit guarantees (i.e. government-backed insurance against bad debts for overseas sales), financial help (such as government grants to the aircraft or shipbuilding industry) and state assistance via the Foreign Office;
- (b) for imports -complex import regulations and documentation, or special safety standards demanded from imported goods and so on.

When a government gives grants to its domestic producers, e.g. regional development grants for new investments in certain areas of the country, or grants to investments in new industries, the effect of these grants is to make unit production costs lower. These give the domestic producer a cost advantage over foreign producers in export markets as well as domestic markets.

Government action to devalue or depreciate the currency

If a government allows its currency to fall in value, imports will become more expensive to buy.

This will reduce imports by means of the price mechanism, especially if the demand and supply curves for the products are elastic. For example, if the exchange rate between sterling and the US dollar is £1 = \$1.60, a good imported from the USA to the UK at a cost of \$8,000 would cost the UK buyer £5,000. Now if the government takes action to reduce interest rates, say, which has the effect of weakening the value of sterling, the exchange rate might change to £1 = \$1.50. The same good costing \$8,000 will now cost a UK buyer £5,333 i.e. £333 more than before. At this higher price, the total UK demand for the US good will probably fall.

The extent of the fall in imports will depend on the price elasticity of demand in the UK for the US good.

The arguments in favour of protection

The arguments for protection are as follows.

- (a) Protectionist measures can be taken against imports of cheap goods that compete with higher-priced domestically-produced goods, and so preserve output and employment in domestic industries.
- (b) Measures might be necessary to counter 'dumping' of surplus production by other countries at an uneconomically low price.

- (c) Protectionist measures by one country are often implemented in retaliation against measures taken by another country that are thought to be unfair.
- (d) There is an argument that protectionism is necessary, at least in the short term, to protect a country's infant industries that have not yet developed to the size where they can compete in international markets.
- (e) Protection might also help a country in the short term to deal with the problems of a declining industry.
- (f) Protection is often seen as a means for a country to reduce its balance of trade deficit, by imposing tariffs or quotas on imports.

The arguments against protection

The arguments against protection are as follows.

- (a) Because protectionist measures taken by one country will almost inevitably provoke retaliation by others, protection will reduce the volume of international trade. This means that the benefits of international trade will be reduced.
- (b) Because of retaliation by other countries, protectionist measures to reverse a balance of trade deficit are unlikely to succeed. Imports might be reduced, but so too would exports.
- (c) It is generally argued that widespread protection will damage the prospects for economic growth amongst the countries of the world, and protectionist measures ought to be restricted to 'special cases' which might be discussed and negotiated with other countries.
- (d) Protection creates political ill-will amongst countries of the world and so there are political disadvantages in a policy of protection.

As an alternative to protection, a country can try to stimulate its export competitiveness by making efforts to improve the productivity and lower the costs of domestic industries, thus making them more competitive against foreign producers.

Hidden subsidies and exchange rate devaluation or depreciation are examples of indirect protectionist measures, but other measures, such as funding industrial training schemes and educational policies, might in the longer term result in improvements in domestic productivity.

General Agreement on Tariffs and Trade (GATT)

This agreement was signed by 23 countries in 1947 as an attempt to promote free trade. The aims of GATT are:

- (a) to reduce existing barriers to free trade;
- (b) to eliminate discrimination in international trade;
- (c) to prevent the growth of protection by getting member countries to consult with others before taking any protectionist measures.

The 'most-favoured nation' principle applies whereby one country (which is a member of GATT) which offers a reduction in tariffs to another country must offer the same reduction to all other member countries of GATT.

GATT now consists of about 100 members.

GATT has succeeded in reducing world tariffs, but there are serious problems which can arise.

- (a) A country wishing to join GA TT must consider the effect of reducing tariffs on its balance of payments and domestic economy.
- (b) Special circumstances (e.g. economic crises, the protection of an infant industry, the rules of the EC) may be admitted whereby protection or special low tariffs between a group of countries are allowed.
- (c) A country in GA TT may prefer not to offer a tariff reduction to another country because it would have to offer the same reduction to all other GA TT members.
- (d) Protectionist measures (and the threat of further protection) are widespread, and GA TT has serious difficulties in persuading member countries to remove them

The EC and the single European market 1992

The European Community (EC) constitutes a localised free trade agreement among member countries incorporating protectionist measures against other trading nations.

The end of 1992 marked the removal of many physical, technical and fiscal barriers among member states, thus creating a large multinational European 'Single Market'. This objective was embodied in the Single European Act of 1985.

It may be useful to look briefly at what the 1992 changes mean.

- (a) Many of the technical barriers to the free movement of goods between countries in the EC have been removed. Technical barriers fragment the market, and force producers to manufacture modified goods, at higher cost, for each separate fragment. The barriers are created by, for example, differing national product standards for the Quality and safety of goods, differing food laws and differing regulations for controlling the access of pharmaceuticals (medicines) to the market.
- (b) Public purchasing. Public contracts are often lucrative for companies that win them. In the past there has been a tendency for contracts to be awarded by governments to 'national' companies. EC rules will open up the market for goods purchased by governments of member countries, by making sure that all companies in the Community have an equal chance to seek individual public contracts.
- (c) Telecommunications. The EC aims to open up telecommunications markets to greater competition, by eliminating differences in national standards for telecommunications equipment.
- (d) Information technology. The EC is promoting the creation of a common set of standards for IT equipment (OSI standards), so that IT equipment can be linked and work together, regardless of who manufactures each item.
- (e) Financial services. There is a continuing process of abolition of national barriers which restrict the provision of financial services between EC countries.
- (f) Capital movements. The EC seeks to achieve the complete liberalisation of movements of capital within the Community.
- (g) Transport. Measures to liberalise transport services will increase competition and make it easier for companies in one EC country to compete in other EC countries. Liberalisation measures are being applied to road haulage, shipping and civil aviation.
- (h) Professional services, in the past professionally qualified people in one country have usually been restricted from practising in another EC country unless they requalify. It is proposed to remove all such restrictions

The Balance Of Payments. Exchange Rates

Your objectives

After completing this chapter you should:

- (a) understand what is meant by the balance of payments and be aware of the components in its make-up;
- (b) be aware of the effects of surpluses or deficits in the balance of payments
- (c) understand the meaning and the effects of a country's terms of trade
- (d) be aware of the importance of exchange rates and understand how they may be determined

The balance of payments

The nature of the balance of payments

The balance of payments is a statistical 'accounting' record of a country's international trade transactions (the purchase and sale of goods and services) and capital transactions (the acquisition and disposal of assets and liabilities) with other countries during a period of time.

Under the current method of presentation of the UK balance of payments statistics the broad classifications of transactions are:

- (a) current account transactions. These are sub-divided into transactions in 'visibles' (i.e. goods) and transactions in 'invisibles' (i.e. services). The current account is therefore the record of a country's trading in exports and imports of goods and services;
- (b) changes in the UK's external assets and liabilities.

The sum of the balance of payments accounts is zero

The sum of the balance of payments accounts must always be zero (ignoring statistical errors in collecting the figures).

This is because every transaction in international trade has a double aspect. In the balance of payments, every 'plus' item should have a matching 'minus' item.

If the balance of payments in principle sums to zero, you may wonder what is meant by a surplus or deficit on the balance of payments. When journalists or economists speak of the balance of payments they are usually referring to the deficit or surplus on the current account, or possibly to the surplus or deficit on visibles only. To appreciate these terms, it is helpful to look at balance of payments items in a bit more detail. Below, we use UK statistics to illustrate principles which apply equally to the balance of payments of every other country too.

Components of the balance of payments

The current account

We will now consider the components of the balance of payments in more detail, starting with the current account. The visible and invisible sections of the balance of payments together make up the balance on current account.

The visible balance

The visible balance is sometimes referred to as the 'balance of trade'. The visible balance is the difference between the value of exported goods from the and imported goods to the.

The invisible balance

The invisible balance consists of services, interest, profit and dividends, and transfers. The annual invisible balance is invariably in surplus for the UK. The net surplus on financial services is a major reason for this.

Services include transport (by sea and air, both passenger and cargo), tourism, financial (banking, insurance, brokerage, etc) and government services (chiefly due to military and diplomatic presence overseas).

These are important ingredients in our invisibles section of the account and have grown in importance in recent years, particularly inward tourism and earnings from financial services.

Interest, profits and dividends consist of items such as:

- (a) direct investment earnings. These are the share of profits in overseas branches, overseas subsidiary companies and overseas associated companies. Direct investment earnings might bring income into the country (e.g. the profits of UK banks operating overseas) or cause outflows (profits of overseas firms investing in the UK);
- (b) portfolio investment earnings -i.e. interest and dividends on stocks and shares held in overseas securities by UK residents, or held in UK securities by overseas residents;
- (c) interest on borrowing and lending abroad by UK banks

UK external assets and liabilities

Transactions in UK external assets and liabilities record the increases or decreases in:

- (a) foreign assets, including foreign currency, that are held by UK residents (including the UK government);
- (b) the liabilities of UK to residents of other countries. These are increases or decreases in UK assets, including sterling, that are held by individuals, firms or governments in other countries.

Transactions in official reserves

An item under the heading of (transactions in) assets in the balance of payments is 'drawings on or additions to the official reserves'.

The official reserves consist of mainly gold and convertible foreign currencies, held in the government's Exchange Equalisation account with the Bank of England. (Other countries similarly have official reserves, which are kept and managed on behalf of the government by the central bank.) Governments use these reserves to balance the overseas accounts and also to deal in the foreign exchange markets.

In past years, movements on the official reserves were given greater prominence in the UK balance of payments statistics, but other international capital transactions are so large that they now 'dwarf' any changes in official reserves, and so changes in the official reserves are no longer given the same prominence as before. Instead, they are included as just another item in the assets section of the balance of payments account.

Political pressure might build up within the importing countries to impose tariffs or import Quotas.

It might therefore be argued that a country has a good balance of payments position if, in the long run, it has neither surplus nor deficit on its current account, or at least if, in the long run, its current account deficit is matched by inflows of direct investment (ie long-term) capital.

CAUSES AND CONSEQUENCES OF A TRADE DEFICIT

Causes

- Poor exports caused by o a high exchange
 - rate keeping prices of export goods in foreign markets too high
 - inefficient production, keeping costs of output too high and uncompetitive in foreign markets
- Excessive imports consumer goods
- Imports of raw materials and capital goods by firms, seeking to expand their own output

Consequences

- The country is importing more goods than it exports, and so the country's general standard of living is being improved
- The country must import capital from abroad. Foreign investors are only likely to invest
 - for high interest rates/high profits and
 - if they still have confidence in the country's economic prospects
- Unless sufficient foreign investment is forthcoming, there will be pressures on the country's currency to depreciate, because of excess supply over demand for the currency in the foreign exchange markets

Practical ways of rectifying a current account deficit

The government of a country with a balance of payments deficit will usually be expected to take measures to reduce or eliminate the deficit. A deficit on current account may be rectified by one or more of the following measures:

- (a) a depreciation or devaluation of the currency;
- (b) direct measures to restrict imports, such as tariffs or import quotas or exchange control regulations;
- (c) domestic 'deflation' to reduce aggregate demand in the domestic economy.

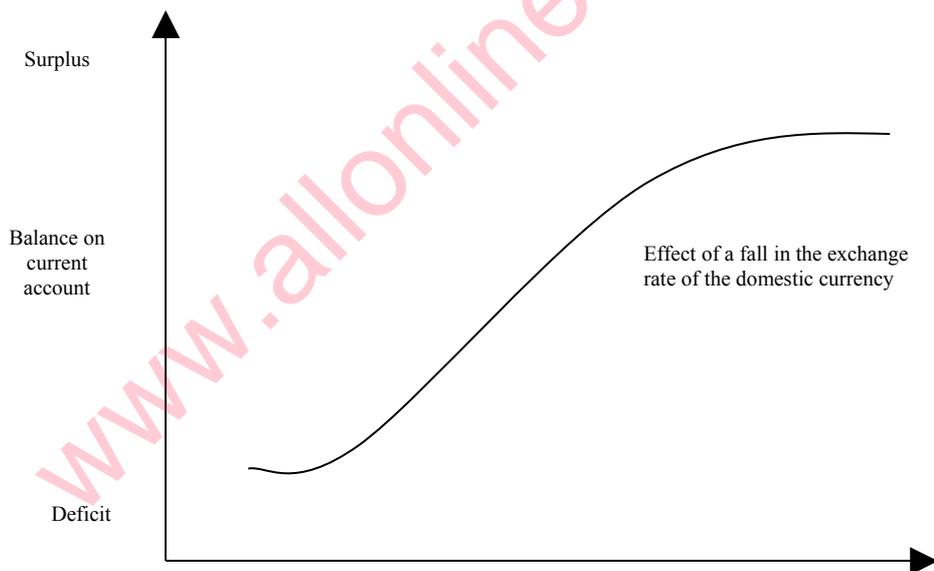
If a country imports raw materials and exports manufactured goods which are made with those materials, the cost of imported raw materials will rise, and so producers will have to put up their prices to cover their higher costs. There will be a net fall in export prices, as explained above, but perhaps not by much.

Because the effect of depreciation or devaluation depends on price elasticities of demand in this way, it might be the case that depreciation of the currency on its own would be insufficient to rectify the balance of payments deficit, unless an extremely large depreciation took place.

The effects of a fall in the exchange rate

We can now consider what would happen to the UK current account balance if there were a fall in the value of sterling.

- (a) The immediate effects will depend on the elasticity of demand for imports. In the short run, demand is likely to be fairly inelastic and so total expenditure on imports will rise.
- (b) Exports will be cheaper in overseas markets (in foreign currency), but in the short run UK exporters might be unable to increase their output to meet the higher demand.
- (c) Until UK industry adjusts to the change and increases its output of exported goods and home-produced substitutes for imported goods there will be a deterioration in the current account of the balance of payments.
- (d) After a time lag, production of exports and import substitutes will rise, so that:
 - (i) the volume of exports will rise, thereby increasing the sterling value of exports (regardless of sterling's lower exchange rate);
 - (ii) the volume of imports will fall further. This will improve the current account balance. The extent of this improvement will depend on the price elasticity of demand for UK exports abroad and the price elasticity of demand for foreign imports in the UK.
- (e) The improvement in the balance of payments will have some limit, and the current balance should eventually level off. The effect of the falling exchange rate on the current balance can be shown by a J' curve (Figure 1).



The Terms of Trade

Introduction

The balance of trade for any country depends on two things

- (a) the volume of goods exported and imported and
- (b) the relative prices of exports and imports

The terms of trade can be defined as the quantities of domestic goods that a country must give up to obtain a unit of imported goods.

- (a) If circumstances change so that X must produce and export 1.5 units of its goods to obtain 1 unit of goods from Y, the terms of trade will have shifted against X and in favour of Y.
- (b) If circumstances change so that X must produce and export only 1 unit of its goods to obtain 1 unit of goods from Y, the terms of trade will have shifted in favour of X and against Y.

The ratio of export to import prices (the terms of trade) determines the volume of exports necessary to pay for a given volume of imports or, meaning the same thing, the volume of imports that can be purchased with the proceeds of a given volume of exports.

Other things being equal, if the price of exports falls relative to that of imports (a fall in the terms of trade) the trade balance will deteriorate, or vice versa.

Note that the trade balance depends not just on the physical volume of exports and imports, but also on the prices at which they are traded.

Measuring the terms of trade and changes in the terms of trade

The terms of trade, are measured as:

$$\frac{\text{unit value of exports}}{\text{unit value of imports}}$$

In practice economists are usually concerned, not with a measurable value for the terms of trade, but with a measure of changes in the terms of trade, (e.g. from one year to the next).

Using indices for the average prices of imports and exports, the movement in the terms of trade between 1992 and 1993 would be computed as:

$$\frac{\text{Price of exports 1993/price of exports 1992}}{\text{Price of imports 1993/price of imports 1992}}$$

Change in a country's terms of trade occur because of:

- (a) a change in the composition of exports or imports. In the UK two main things (fewer oil exports and manufacturers trading up to higher-price products for export) have improved the UK's terms of trade since 1985;
- (b) lower or higher prices of imports/exports (e.g. the oil price collapse in 1985, which worsened the terms of trade for the UK).

A government has limited 'powers' to influence its country's terms of trade, since it cannot directly influence the composition nor the prices of imports and exports although it can affect the terms of trade through a revaluation or devaluation of the currency which would alter relative import/export prices.

What does a change in the terms of trade mean ?

If a country's terms of trade worsen, the unit value of its imports will rise by a bigger percentage than the unit value of its exports. The terms of trade will worsen when the exchange rate of the currency depreciates in value against other currencies.

If a country's terms of trade improve, the unit value of its exports will rise by a bigger percentage than the unit value of its imports. The terms of trade will improve when the exchange rate of the country's currency appreciates in value against other currencies.

It would seem logical to assume that an improving terms of trade is 'good' for a country and a worsening terms of trade is 'bad'. But this is not necessarily the case.

What is the effect of a change in the terms of trade? This should be considered in the context of the country's balance of payments.

The effect of changes in the terms of trade

If the terms of trade worsen for a country, the country will be unable to afford the same volume of imports, or else its balance of payment position will deteriorate. In contrast, a country with improving terms of trade will be able to afford more imports or will improve its balance of payments.

Changes in the terms of trade affect a country's balance of payments via the price elasticity of demand for the goods traded. If a country's terms of trade improve, so that the price of its exported goods rises relative to the price of its imported goods, there will be a relative fall in the volume of goods exported and a rise in the volume of imports. The size of this fall in exports and increase in imports will depend on the price elasticities of demand for exported goods in foreign markets and imported goods in the country's domestic markets.

Exchange rates

We have already mentioned exchanged rates several times. In simple terms, the concept of an exchange rate is familiar to anyone who has ever taken a day trip across the Channel. But in this section we look at the importance of exchange rates in some detail.

An exchange rate is the rate at which one country's currency can be traded in exchange for another country's currency.

Although it is convenient to refer to the 'exchange rate' for currency -e.g. the exchange rate for sterling -every traded currency in fact has many exchange rates. There is an exchange rate with every other traded currency on the foreign exchange markets, so that there is an exchange rate for sterling with the US dollar, the Canadian dollar, the yen, the Deutschmark, the French franc, and so on.

Foreign exchange dealers make their profit by buying currency for less than they pay for it and so there are really two exchange rates, a selling and a buying rate e.g.

	<i>Bank's buying rate</i>	<i>Bank's buying rate</i>
£/US dollar exchange rate	\$1.5020	\$1.5080

When exchange rates are quoted in the press, the 'middle rate' between the selling and buying rates would be used. In the example above, the sterling/US dollar exchange rate would be mid-way between \$1.5020 and \$1.5080, i.e. \$1.5050 to £1.

Spot rates and forward rates

Broadly speaking, there are two ways in which foreign currency is bought and sold:

- (a) spot -i.e. for immediate 'delivery'.
- (b) forward- i.e. for delivery at a date in the future.

Thus, a UK firm might receive US\$100,000 from a US customer, and sell it 'spot' to a bank, to receive sterling immediately (in practice three days after the contract is made). If the exchange rate is \$1.8000 to £ 1, the UK firm would receive £55,555.56.

If a firm knows that it is going to receive some foreign currency in the near future, which it will want to sell in exchange for domestic currency, it can make a forward exchange contract with a bank, at an exchange rate that is specified in the contract. Thus, if a firm knows that it is going to receive US\$100,000 in three months' time, it can make a forward exchange contract 'now' to sell the US dollars in three months' time at a specified exchange rate. If the 'spot' rate is \$1.8000 to £1, the 'forward' rate may be higher or lower than \$1.8000 (depending on comparative interest rates in the USA and the UK).

Factors influencing the exchange rate for a currency

The exchange rate between two currencies -ie the buying and selling rates, both 'spot' and forward -is determined primarily by supply and demand in the foreign exchange markets. Demand comes from individuals, firms and governments who want to buy a currency and supply comes from those who want to sell it.

Supply and demand in turn are influenced by'

- (a) the rate of inflation, compared with the rate of inflation in other countries; interest rates, compared with
- (b) interest rates in other countries;
- (c) the balance of payments;
- (d) speculation;
- (e) government policy on intervention to influence the exchange rate.

Purchasing power parity theory

If the rate of inflation is higher in one country than in another country, the value of its currency will tend to weaken against the other country's currency.

Purchasing power parity theory, which developed in the 1920s, attempted to explain changes in the exchange rate exclusively by the rate of inflation in different countries. The theory predicts that the exchange value of a foreign currency depends on the relative purchasing power of each currency in its own country. As a simple example, suppose that there is only one commodity, which costs £110 in the UK and 880 francs in France. The exchange rate would be £1 = 8 francs. If, as a result of inflation, the cost of the commodity in the UK rises to £120, the exchange rate would adjust to:

$$(8 \times \frac{110}{120}) \times \text{£}1 = 7.33 \text{ francs}$$

If the exchange rate remained at £1 = 8 francs, it would be cheaper to import more of the commodity from France for £110 and the UK would have a balance of trade deficit. This would only be corrected by an alteration in the exchange rate, with the £ weakening against the franc.

Purchasing power parity theory states that an exchange rate varies according to relative price changes, so that

$$\text{'old' exchange rate} \times \frac{\text{price level in country A}}{\text{price level in country B}} = \text{'new' exchange rate}$$

The theory was soon found to be inadequate to explain movements in exchange rates in the short term, mainly because it ignores payments between countries (i.e. demand and supply transactions) and the influence of supply and demand for currency on exchange rates.

Interest rates and the exchange rate

It would seem logical to assume that if one country raises its interest rates, it will become more profitable to invest in that country, and so an increase in (mainly short- term) investment from overseas will push up the exchange rate because of the extra demand for the currency from overseas investors.

This is true, but there is a limit to the amount of investment capital that will flow into a country because of higher interest rates. A major reason this is that investors may expect a 'risk premium' for investing in a high interest rate currency if they fear that the currency will depreciate in value.

The balance of payments and the exchange rate

Purchasing power parity theory is more likely to have some validity in the long run, and it is certainly true that the currency of a country which has a much higher rate of inflation than other countries will weaken on the foreign exchange market. In other words, the rate of inflation relative to other countries is certainly a factor which influences the exchange rate.

Although this influence is obvious, it is not predominant. This is apparent from the fact that if exchange rates did respond to demand and supply for current account items, then the balance of payments on the current account of all countries would tend towards equilibrium. This is not so, and in practice other factors influence exchange rates more strongly.

Government intervention

The government can intervene in the foreign exchange markets:

- (a) to sell its own domestic currency in exchange for foreign currencies, when it wants to keep down the exchange rate of its domestic currency. The foreign currencies it buys can be added to the official reserves;
- (b) to buy its own domestic currency and pay for it with the foreign currencies in its official reserves. It will do this when it wants to keep up the exchange rate when market forces are pushing it down.

The government can also intervene indirectly, by changing domestic interest rates, and so either attracting or discouraging investors in financial investments which are denominated in the domestic currency. Purchases and sales of foreign investments create a demand and supply of the currency in the FX markets, and so changes in domestic interest rates are likely to cause a change in the exchange rate.

Exchange rate policies of governments

Exchange rate policy options

We shall now go on to consider in more detail the different exchange rate policies which are open to governments. These may be categorised as:

- (a) fixed exchange rates;
- (b) free floating exchange rates;
- (c) margins around a movable peg;
- (d) managed floating.

Fixed exchange rates

A policy of rigidly fixed exchange rates means that the government of every country in the international monetary system must use its official reserves to create an exact match between supply and demand for its currency in the FX markets, in order to keep the exchange rate unchanged. Using the official reserves will therefore cancel out a surplus or deficit on the current account and non-official capital transactions in their balance of payments. A balance of payments surplus would call for an addition to the official reserves, and a deficit calls for drawings on official reserves.

The official reserves could in theory consist of any foreign currency (or gold) within the fixed exchange rate agreement. The exchange rates of the various currencies in the system might all be fixed against each other. However, for simplicity and convenience, it is more appropriate to fix the exchange rate for every currency against a standard. The standard might be:

- (a) gold. If every currency is valued in terms of gold, official reserves would consist mainly, or even entirely, of gold;
- (b) a major currency, such as the US dollar;
- (c) a basket of major trading currencies. For example, as explained below, the ecu (European Currency Unit) is an 'international currency' which is based on a 'basket' of currencies of EC countries.

A fixed exchange rate system removes exchange rate uncertainty and so encourages international trade. It also imposes economic disciplines on countries in deficit (or surplus). However, this restricts independence of domestic economic policies. A government might be forced to keep interest rates high or to reduce demand in the domestic economy (e.g. by raising taxes and so cutting the demand for imports) in order to maintain a currency's exchange rate and avoid a devaluation.

If levels of inflation differ widely in countries subscribing to a fixed exchange rate regime, the regime may not survive for long. The high inflation countries will be forced to devalue in order to keep their exports competitive and to reduce imports.

Free floating exchange rates

Free floating exchange rates are at the opposite end of the spectrum to rigidly fixed rates. Exchange rates are left to the free play of market forces and there is no official financing at all. There is no need for the government to hold any official reserves, because it will not want to use them.

Floating exchange rate systems (free floating and managed floating) have been criticised in the past because they allow wide fluctuations in exchange rates. Certainly, in the foreign exchange markets

today, there are large fluctuations which are unsettling for international trade. Floating exchange rates are the only option available to governments when other systems break down and fail.

A movable peg or adjustable peg system

A movable or adjustable peg system is a system of fixed exchange rates, but with a provision for:

- (a) the devaluation of a currency, e.g. when the country has a persistent balance of payments deficit;
- (b) the revaluation of a currency, e.g. when the country has a persistent balance of payments surplus.

Margins around a movable peg

A movable peg system provides some flexibility. Exchange rates, although fixed, are not rigidly fixed, because adjustments are permitted. Even so, it is still fairly inflexible, because governments only have the choice between a revaluation/devaluation or holding the exchange rate steady. A more flexible system would allow some minor variations in exchange rates. For example, the exchange rate between sterling and the US dollar might be fixed at \$2 to £1, but governments might only be required to maintain the exchange rate within a margin of, say, 2% on either side of this rate. If this were the case, the UK government would undertake to keep the exchange rate for sterling between \$1.96 and \$2.04 to £1. However, if the UK were to run into a fundamental balance of payments disequilibrium, a devaluation (or revaluation) of sterling would occur and the UK government would then undertake to maintain exchange rates within the required margins of the new exchange rate.

The European exchange rate mechanism (ERM) is an adjustable peg system

European monetary cooperation

The European Monetary System (EMS)

The European Community opted for a 'local' international agreement on exchange rates, originally known as the European Snake but amended in 1979 to the EMS system. In the EMS:

- (a) there is a scheme of margins around a central peg for exchange rates between the currencies of the countries in the EMS; but
- (b) a policy of managed floating between their currencies and the currencies of countries outside the system.

The Exchange Rate Mechanism (ERM) is the exchange rate agreement of the European Monetary System, formed on 13 March 1979. The United Kingdom has been a member of the EMS since then, but did not join the exchange rate mechanism of the EMS until October 1990 and left the mechanism in 1992.

The main features of the EMS are as follows.

- (a) It provides for the system of exchange rates for member currencies commonly known as the 'ERM'. Each currency has a 'central parity' rate against the ecu.
- (b) The EMS created the new currency, the European Currency Unit or ecu. An ecu is a unit of currency based on a 'basket' of the currencies of the participating countries. The value of the ecu therefore depends on the weightings given to each individual currency in the basket, and these weightings were based on the relative importance of each currency in European Community trade at the time the ecu was devised.
- (c) Within the ERM the exchange rate of the currency of each member country is permitted to vary within a margin of plus or minus a specified percentage against its central parity. Each currency must also keep within the same limit of each other currency within the system.
- (d) The exchange rate of each currency is therefore pegged against the ecu and also against every other ERM member currency.

- (e) As the ERM is a 'movable peg' system, occasional revaluations or devaluations of the central parity rates can occur, but only as a last resort measure.

www.allonlinefree.com