Edinburgh Napier UNIVERSITY

School of Engineering & the Built Environment

CTR Postgraduate Programmes

CTR11116

Transport Economics and Appraisal

Study Guide 1 – Transport Economics

Author:

Jonathan Cowie (2004)

All Units Updated November 2009

Published by:

Edinburgh Napier University
School of Engineering and the Built Environment
10 Colinton Road
Edinburgh
EH10 5DT
Scotland
United Kingdom

TRANSPORT ECONOMICS STUDY GUIDE ONE

Introduction

Welcome to the Transport Economics and Appraisal module. This is the first of thirteen units, which will provide an introduction to the module by outlining the dynamics of the transport market. The first six units in the Transport Economics and Appraisal module are on the subject of Transport Economics, although note that we often run this module the other way around, i.e. Transport Appraisal first. These units are:

- Introduction: Dynamics of the Transport Market
- Consumers: Elasticity of Demand
- Characteristics of Public Transport Operators
- Transport Costs incurred by Public Transport Operators
- Pricing Policy and Market Segmentation (supplementary)
- Transport Economic Regulation
- Transport Subsidy

At the outset, I should reassure if you are worried about the level of economics knowledge within these units on Transport Economics. I will take an approach of providing simple economic definitions and concepts, and then apply them to transport situations. Take it at your own speed so that you can understand the economic concepts fully before we apply them to a number of transport case studies. The approach that I have taken however is not to fill the pages with needless woffle, but rather to examine the key issues. You should therefore read through each of the units more than once, as second and third readings may produce more of an understanding of the topics under consideration. Furthermore, as you go through the units you should consider journeys that you have made and relate them to the course.

This unit, the first, provides an introduction to Transport Economics. Units 1 and 2 focus on individual consumers and their demand for transport services. It will examine how demand varies, with analysis of their elasticity of demand. The focus will then turn to the supply side of the market. Unit 3 will examine cost and productivity as the level of output varies. The economic concepts of the margin and profit will also be considered. Unit 4 investigates market structure under the title of the theory of the firm. Unit 5 is supplementary and takes the issues in Unit 4 further to look at the pricing policy of public transport operators and the market segmentation that exists. It will examine why people often pay different prices for the same public transport journey. Unit 6 focuses on a key area in the provision of (public) transport services, that of economic regulation. Finally Unit 7 focuses on transport subsidy, with a particular emphasis on the rationale for providing subsidy to public transport operators in the UK.

RECOMMENDED READING

For fairly obvious reasons, the key text and, again for fairly obvious reasons, highly recommended reading is:

Cowie, J. (2010). *The Economics of Transport, a theoretical and applied perspective*. Routledge, Aldershot.

Other key texts on Transport Economics are:

Mallard G. And S. Glaister (2008). Transport Economics, Palgrave MacMillan.

Cole S (2005), *Applied Transport Economics. Policy, Management and Decision Making, Thrid edition,* Kogan Page Ltd, Great Britain.

Button K J (1993), *Transport Economics. Second Edition,* Edward Elgar Publishing Ltd, Cheltenham, UK.

The first text does provide an introductory level outline of the subject, however tends to be quite abstract in its approach. Of the other two, the first tends to be far too general and too simplistic, whilst Button's text, although excellent, is too advanced for an introductory course. Whilst such critical comment may be expected from some who has written a text on transport economics(!), the above comments were the whole motivation for writing it! Additional references however will be given for these texts in the relevant section, but there really is only one book you should look at!

For a more general economics text, you should consult the following:

Sloman J (2009), *Economics, 7th Edition,* Financial Times/Prentice Hall.

An old 6th edition copy (2005) of Sloman's text would also suffice.

LEARNING OUTCOMES

Once you have worked your way through this unit you should be able to:

- Grasp introductory economic concepts which can be applied to the transport sector; and
- Understand the cost components of owning and using a motor car.

A VERY BRIEF PREAMBLE ON THE STUDY OF TRANSPORT ECONOMICS

Transport Economics combines two subjects, transport and economics, applying economic concepts to the transport sector. Economics has been formally defined as the "science of the production and distribution of wealth.". You will note that part of that definition uses the word 'science'. Economics is obviously not one of the natural sciences, but as it is the study of an aspect of human behaviour it is a social science, and therefore fits into the category along

with the other social sciences psychology, sociology, anthropology and so on. The major implication however of it being a social science is that there are no right or wrong answers, as we are dealing with human behaviour and all humans beings are different, although note there will be some answers that will be more 'right' than others. Most analysis therefore is carried out at the aggregate level, because although we may not be able to predict how a particular individual may react we can (or should be able to) predict how 'people' will react.

There is one more preliminary before we begin. The study of Economics involves a large number of diagrams. Diagrams are used to illustrate concepts and ideas. Most texts introduce the idea of the use of diagrams in Economics as 'one picture is worth a thousand words', a sentiment that this author does not agree with. To the average student of Economics, which these days consist of those undertaking some form of business degree in subjects such as Marketing, Human Resource Management, Business Information and so on, the use of diagrams in Economics only produces confusion. Diagrams provide ample opportunities for misunderstandings and misconceptions – after all, once you've seen one diagram you've seen them all! Unfortunately, the use of diagrams in Economics is essential. Why this is the case however is not because 'one (very confused) picture is worth a thousand (rambling) words (of complete rubbish)', but rather that the use of diagrams provides structure and focus in the explanation of Economic concepts. This applies whether that is me explaining them to you (course notes/lectures/seminars) or you explaining them to me (exams/courseworks/dissertations). Diagrams are almost essential to organise thinking and provide a logical and coherent explanation. For example the analysis of many economic situations requires the working through of first principles and the drawing of some basic diagrams in order to understand the current context.

Self Assessment Exercise 1

On an A4 sheet of paper, write down a list of key words that you believe relate to the study of Economics. Keep this in a safe place until you have finished Unit 1, as you will need it to complete one of the exercises at the end of that unit.

In a separate but related exercise, produce a list of what you consider to be the main transport economics issues in the current provision of private, public and freight transport.

UNIT ONE
INTRODUCTION: THE DYNAMICS OF THE TRANSPORT MARKET

LEARNING OUTCOMES

Once you have worked your way through this Unit, you will be able to:

- ◊ Identify the key principles that underpin the study of Economics,
- ◊ Define main characteristics that outline the free, command and mixed market economies
- ◊ Understand the key concepts of demand and supply
- ◊ Illustrate the market mechanism.

We start our study of transport economics with a more full definition of the economics discipline. Most image Economics is to do with money and all things financial, but what the subject is about can be very simply summarised in three simple words – scarcity, choice and opportunity cost. These three ideas are outlined below:

SCARCITY, CHOICE AND OPPORTUNITY COST

Scarcity is a concept that we normally associate with Third World countries, where a lack of rainwater and the subsequent failure of agricultural produce cause famine and droughts. Scarcity however applies to not only 3rd World economies but all economies, whether 3rd world, developing or advanced. In very simple terms because we cannot have everything that we want, then there is a finite limit on resources that the demand for which exceeds – any resource is therefore scarce, perhaps not at the margin or a basic necessity, but it is scarce nevertheless.

If we cannot have all that we want then we have to make choices. This brings in our second concept of choice and put simply, every choice involves a cost, namely the next best alternative that you could have had when you make a choice. This is known as the opportunity cost of that decision. Thus if a particular society does not have sufficient resources to build both a school and a hospital, it must make a choice between the two. If it chooses to build the school then the opportunity cost of the school is the hospital that was not build. Opportunity cost therefore can be formally defined as 'the next best alternative' foregone and is consequently not assessed using financial criteria.

Although we have introduced most of these three concepts using an individual's perspective, exactly the same principles operate at the society level. This is because any economy (country) cannot provide its citizens with all that they want, i.e. there is scarcity. A choice therefore has to be made with regard to three basic questions –

- what to produce?
- how to produce it?
- and for whom to produce it?

These questions are key to how society organises its resources of production and the system of distribution that will be used. This has led to the development of different economic systems to answer the three questions given above, the two most extreme of which are the command economy and the free market economy.

A command economy is where the state, i.e. the government, directly addresses the three questions posed above. That is, the government decide what to produce, how it will be produced and who will receive the resultant output. In the past this has normally been centred around a system of plans, in which five year plans are subdivided into one year plans, then area plans of production, then by town, by company, by individual plant and so on down. The government also decides how the *factors of production* are employed. Factors of production are the resources that are used in the production process. All production processes can be broken down into requiring three factors of production – land/raw materials, labour and capital. Land/raw materials and labour are fairly straightforward, capital on the other hand is any equipment that is used in the production process. Under a command economy system, the state organises the factors of production to resolve the first two questions of what and how to produce, and distributes the resultant production on the basis of equity i.e. if you work hard you reap the rewards. Note that in theory under such a system there is no need for any form of money, as goods and services are distributed on the basis of decisions by government or some other delegate central body. Such a system was in the past associated with the former communist countries in Eastern Europe, however it is worth highlighting that such systems were not always associated with the political left, as the National Socialist German state was based upon a command economy.

Although the relevance of studying such systems may appear to have disappeared with the collapse of the European communist states (and their subsequent command economies) in the late 1980s and early 1990s, the relevance is still very much applicable today, if only from an important theoretical perspective. Command economies highlight the role that government plays in the running of the economy, and as we will see, this is particularly relevant in the study of transport markets.

At the other end of the spectrum there is the free market economy. At the most extreme example, the government has no input into the decisions of what, how and for whom to produce, as these are left purely to the market to decide. The market therefore will decide the questions of what, how and for whom to produce through the price mechanism and the profit motive. The price mechanism is the device that is used to transmit signals from the market to the various interested parties, which in a strictly free market economy is only made up of buyers and sellers. The whole system is based upon trade and the notion that trade is never a zero sum game, as both parties will (usually) benefit in any exchange. A final important concept of the free market is the idea of consumer sovereignty i.e. the consumer is king. In simple terms, if consumers want more of something they will go out and buy it, this will cause the price of that commodity to rise, and this will send a signal through the price mechanism to producers that consumer want more of that particular product. They will therefore produce more of it.

This allows us to answer the first two of the three questions posed of what, how and for whom to produce. In the free market economy, consumers through the idea of consumer

sovereignty decide what is produced. Producers decide how these goods/services are produced through combining the factors of production in the lowest cost combination in order to obtain the maximum profit. This again is based upon the price mechanism and the 'price' of the various factors of production. This only leaves the question of 'for whom'? In a free market economy, again this is based (in theory!) on equity. If you work hard, you will gain the rewards. Unlike in the command system however, where it is the state that decides the merit of your claim, in a free market economy it is through the price mechanism. If a particular individual possesses highly sought after skills that are in very short supply, then that individual will command a relatively high price in the market place, i.e. high wages. The subsequent accumulation of higher wages will enable that individual to obtain more of the (scarce) resources of the economic system than the average individual. Through the price mechanism therefore, in this case relating to labour markets, the free market systems resolves the question of for whom to produce.

The last of the economic systems we will examine is the mixed market economy. If you have understood the previous two examples then the mixed market economy is the most straightforward of the three. As the name clearly indicates, it is a market based system, i.e. it is based upon prices, and uses a mix of the state and the market to determine the outcome to the questions of what, how and for whom to produce. A simple way to consider the mixed market system is to think of the free market economy, but rather than simply having two agents in the market place, buyers and sellers, there are three, buyers, sellers and the state.

In practice all economies fall into this last category, with the only difference being the level of state involvement in the market. Some economies, such as France and Sweden tend towards a relatively high level of state involvement, whilst others, such as USA and Honk Kong, have a relatively low level of state involvement. As market based systems dominate throughout the globe however, we will virtually spend the rest of our time in this unit studying the workings of the market economy. Before we leave economic systems however, we consider a brief practical example. The example used is public transport services in Glasgow, namely the bus and the train.

PRIVATE AND PUBLIC SECTOR ROLES IN THE PROVISION OF PUBLIC TRANSPORT IN AND AROUND THE GLASGOW CONURBATION

The market for bus services and the market for train services, whilst both operating through the price mechanism, are actually quite different in nature and form. The former tends to operate along the lines of the free market system, whilst the latter tends to operate along the lines of the command system. In the bus market, privately owned companies decide which bus services they will operate, and this is usually based upon providing a network of services that they believe will make them a profitable return i.e. that consumers are willing to pay for. Thus if there should be a reduction in the number of users of a particular bus service, the revenue (amount taken through the fare box) associated with that bus route will fall. If this trend was to continue, there may come a point where the revenue gained from the service fails to cover the cost of operating the service, and hence the service may be withdrawn.

In simple terms, through the market 'consumers' have effectively decided that they no longer want that bus service and the operator, now making losses, may withdraw that service. More radically, if that is the only bus service that the company operates, then the company is liable to go bust. These 'resources' i.e. land, labour and capital, that were employed by the company are then 'freed' to be engaged in producing something more profitable i.e. something that society, through the market system, has decreed that it wants. That's the theory at least but does not always work out that way in practice!

Rail services on the other hand are quite different in nature. Rather than private companies deciding what services to provide it is the state, through the offices of Transport Scotland that makes such decisions. Transport Scotland determines the pattern of rail services to be provided in and around the Glasgow area. They then contract a private sector operator (First Scotrail), whom they pay, to provide these services to their specified pattern and their specified fare structure. As above, if the revenue from a particular service was to fall, then Transport Scotland would have to decide whether or not it wished to provide that service at the now a higher cost. It would make this decision on the basis of whether or not it would be better to withdraw that service and spend the money on some other transport service. Unlike bus services however, the service would not automatically be withdrawn.

We will no doubt return to bus and rail services in future studies, however now we will begin to examine the market economy system in more detail.

THE TRANSPORT MARKET

We will be looking at the dynamics of a transport market in the rest of this unit, and although we have already encountered the term of 'a market', it is useful to formally define what a market is. In simple terms, a market is a meeting place for buying and selling. A simple illustration is the traditional market place in towns with a number of self-assembled stalls selling a variety of goods.

The economic analysis of a market identifies the distinct notions of demand (from buyers) and supply (of sellers) of goods. These terms are described in turn in the following paragraphs. It is worthwhile before we begin however to also formally define 'a good'. A good is an "item bought and sold for gain". The good could be a watch, a hamburger or a motor car. Goods can be labelled in various ways. Our analysis however will mainly revolve around what is known as a *normal good*. That is a good whose demand rises as people's income rises. This is true for almost all goods. The classic example in transport is the motor car. Most people will buy a car as soon as they can afford it. Contrast this with an *inferior good*, which is a good whose demand falls as peoples' income rises. A common example used in transport is bus travel. As over time incomes in real terms have risen considerably, numbers using bus services have fallen dramatically. This is a situation in certain parts of the country that is changing, however that issue is for a later unit.

THE THEORY OF DEMAND

The demand for a good is the "number of units per unit of time that consumers purchase at any given price". Notice that demand relates to actually purchases, not wants as such. Demand is a result of consumers expressing their own preferences between goods at the relative prices that they face, taking into account all money available to them. Demand can be used to describe an individual's preferences, the preferences of a particular market segment or the whole market. Unit 2 examines demand further, with analysis of elasticity of demand and particular characteristics of transport markets.

As regards the 'basket' of goods and services that consumers acquire, this is said to be determined by consumers seeking to maximise their utility. Utility is an economic concept that in simple terms relates to satisfaction. Thus consumers seek to maximise the satisfaction they derive from the basket of goods and services that they purchase with their (limited) income. The first graph is coming up, known as a demand curve. But before we illustrate it, consider the following blank graph below. We have price on the vertical axis and quantity demanded (i.e. amount purchased) on the horizontal axis. All this graph will attempt to do therefore is to show the relationship between the price of a good and the quantity of that good that is purchased. Take a minute and, given the assumption that consumers seek to maximise utility, graph the relationship between price and quantity demanded. In other words, draw a line on the graph that you think will show the relationship between the price of a good and the quantity demanded.

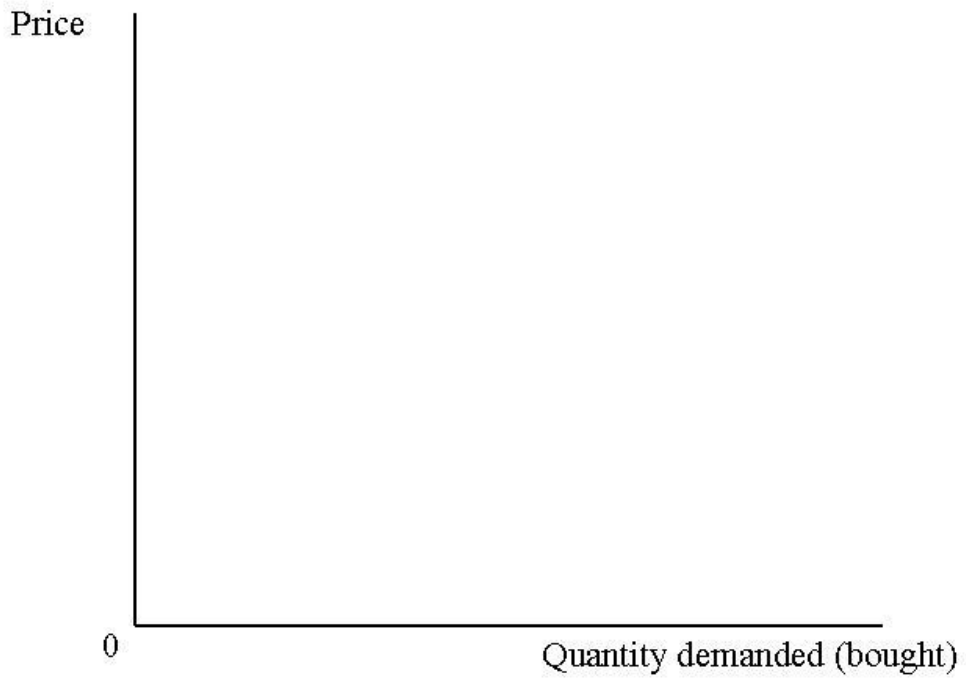


Figure 1.1a: A blank graph for you to show the relationship between price and the quantity demanded for a normal good or service.

Hopefully, if you have followed the discussion, your graph should look something like this:

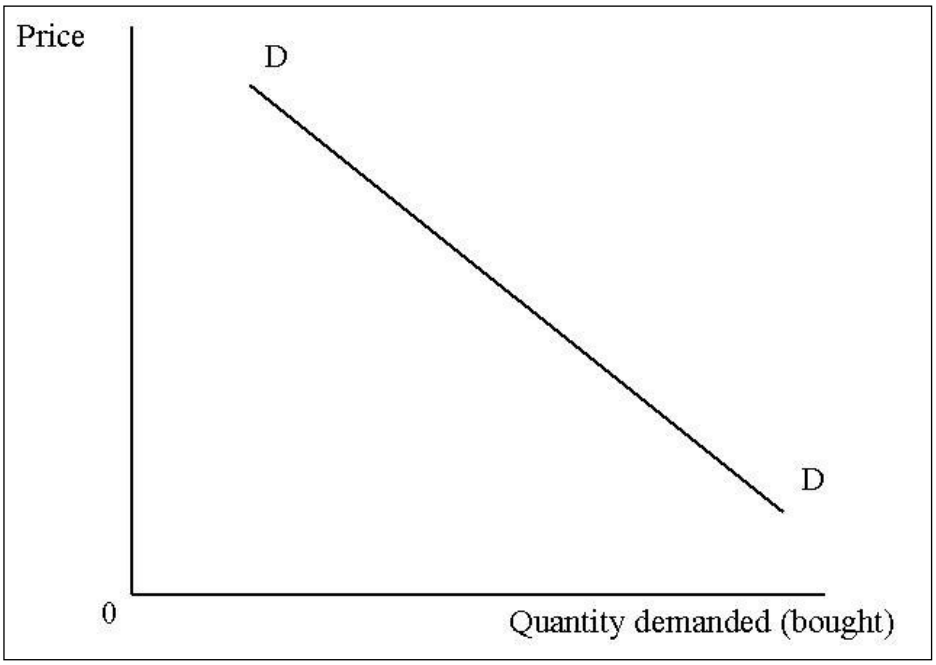


Figure 1.1b: A demand curve

In a sense, although we have used terms like 'quantity demanded' and 'consumers seek to maximise utility', this first graph should be pretty much common sense to most people. As the price of a good or service falls, then the level of (relative) satisfaction arising from that good or service for the money that it cost the consumer to purchase it will increase. Consequently, both existing consumers will purchase more of that good/service and those that could not afford the good or service before can now afford it. With this latter group, as they seek to maximise utility, they believe that its purchase will increase their overall 'satisfaction' (utility) arising out of the basket of goods and services that they purchase from their (finite) income. Therefore, more is sold at lower prices. We have as such our first determinant of demand, the price of the good or service itself. There are however four other basic determinants of demand. These are:

- Income
- The price of other goods or services
- Fashions or trends
- Expectations of future price changes or shortages

For a normal good, any rise in income will cause an increase in demand. We would expect that for each and every price there will be a higher level of demand. The price of other goods and services relates to two specific types of goods and services – substitutes and complements (note the spelling!). Substitute goods are those that could be used in place of the good under analysis. For example, if we were examining the market for butter then an obvious substitute would be low fat spread. If the price of low fat spread was to fall, then we

would expect this to effect the market for butter, and specifically we would expect a decrease in the demand for butter, as low fat spread now represents better value for money. Hence significant numbers of consumers will make the switch. Complementary goods on the other hand is a good that is consumed at the same time. Hence cars and petrol are complementary goods as they are consumed at the same time. If the price of petrol was to rise, we would expect the demand for cars to fall, as what is important is the total price of motoring as such, and hence the overall cost has increased and we would expect demand for cars to fall.

Our third determinant of demand is fashions or trends. At certain points in time, some goods and services will become more fashionable, and hence we would expect that again at each and every price the demand for the good or service would increase. Fashions and trends can be effected by advertising, thus if there was an increase in advertising expenditure we would expect this would affect the demand side of the market and demand would increase. Finally, we have expectations of future price changes or shortages. An expected increase in price in the future would cause an increase in demand in the present, and vice-versa. Similarly, an expected shortage in the future would cause an increase in demand in the present, whilst an expected glut would decrease present demand.

These four determinants are known as the conditions of demand, and as such any change in any of these conditions will affect the relationship between price and the quantity demanded, either causing more to be demanded at each and every price or less. Such changes are represented by a shift in the demand curve, as they change the basic price/quantity relationship. Let us consider a transport example, such as the market for rail services. If the price of an alternative bus service was to fall (i.e. a substitute good), then we would expect that some rail travellers will re-evaluate their use of the train as their mode of transport. This is because the relative value for money of rail travel has fallen in relation to bus travel and they may consequently switch to using the bus. Certainly some of them will do so. This would be represented by a shift in the demand curve (for rail travel) to the left, showing a decrease in the quantity demanded at each and every price. This is shown in figure 1.2.

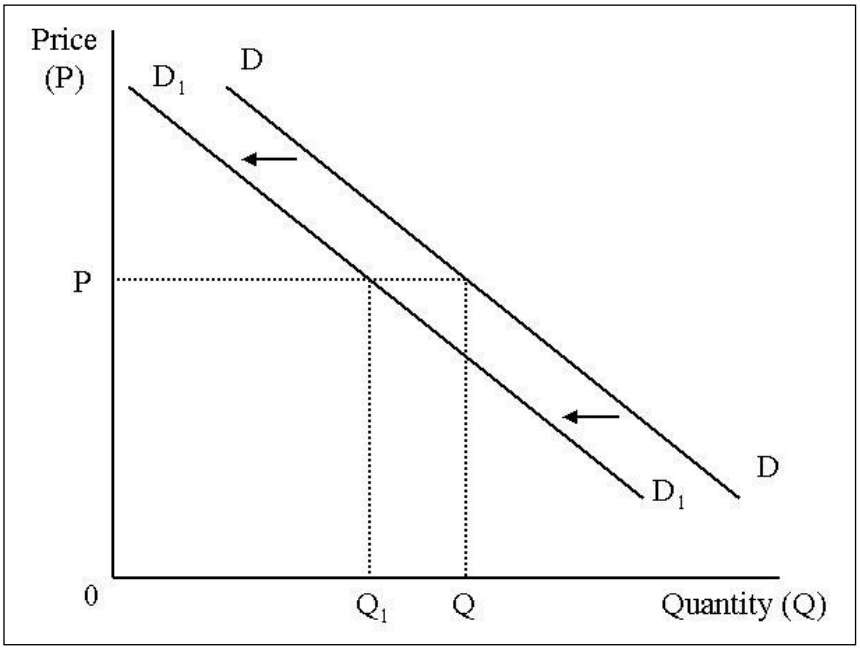


Figure 1.2. Change in the conditions of demand for rail services shown by a shift in the demand curve (decrease in demand)

In figure 1.2, at the original price of P the level of quantity demanded for rail services is given by Q. With the fall in the price of bus services however, less rail services are demanded at each and every price. Hence at the price of P, the level of quantity demanded falls to Q_1. This is the basic change in the price/quantity relationship.

Although we have used a change in the price of a substitute to illustrate shifts in the demand curve, the same logic applies to each of the four determinants of demand introduced above. You should now work through Self Assessment Exercise 1.1 at the end of this unit to make sure that you understand the principles involved.

Further Issues in Demand for Transport

Before we leave demand, there are several important concepts that relate specifically to demand for transport services that need to be highlighted. Firstly, unlike most other products, where they are demanded for their own benefit, most transport is demanded for some other purpose. For example, I need to earn an income (demand for work), but I have to travel in order to access employment. Likewise a large proportion of leisure activities require transport services in order to access. Transport is said therefore to be a derived demand, as the demand for transport is derived from the demand for some other good or service, it is not as such demanded for its own benefit.

Secondly, the consumption of transport services is immediate. If I buy a Mars bar from a shop, I can take it away and eat it several hours later. With most transport services, particularly local services, consumption is immediate. This is problematic because demand is not consistent over a period of time, whether that is related to the time of day, day of the week or time of the year. In particular demand varies according to:

- Time - time of day. Demand is greater just before schools, offices and factories open (8am-9am) and just after they close (5pm-7pm).
- Time - day of week. Demand is often greater during Friday late afternoon as people go away for the weekend.
- Time - seasonal peak. Demand is greater during the Summer school holidays.
- Change in social habits. As an example, in the 1980s holidaymakers started to prefer the Mediterranean to the UK as a holiday destination, but more recently there has been a shift back towards UK breaks with the rise of a second or third holiday/break.
- Changes in population distribution, such as the construction of housing estates on the edge of towns.

These demand patterns mean that there are times of peak and off-peak demand. Examples of maximum or peak demand would include the daily routine of commuters travelling into a major conurbation for work and seasonal habits of holidaymakers travelling during the Summer.

We will return to the demand side of the market in due course, however for now we switch our attention to the issue of supply.

THE THEORY OF SUPPLY

The supply for a good or service can be defined as the "number of units per period of time that producers sell at any given price". Supply is the result of producers examining market opportunities and eventually deciding on a particular market in which to produce a given good or service. Whilst the underlying assumption of consumers was that they sought to maximise utility, the underlying assumption of producers' behaviour is that they seek to maximise profits i.e. to make as much profit as possible. They will thus enter that market in which they believe they can make the most profit. Notice that this must take account of the opportunity cost of that decision. For example, if an individual was to enter the taxi market and earn a profit in the year of twenty thousand pounds, then in effect what should be deducted from that 'profit' is what they could have earned in net wages by working for somebody else. We will return to this issue in later units, however for the moment we will only concentrate on supply to the market.

Take another minute and consider that assumption of firms seeking to maximise profits. If this is the case, then what would you expect the relationship between the price of a good or service and the quantity supplied of it. As before, draw a line on the graph that you think will show this relationship.

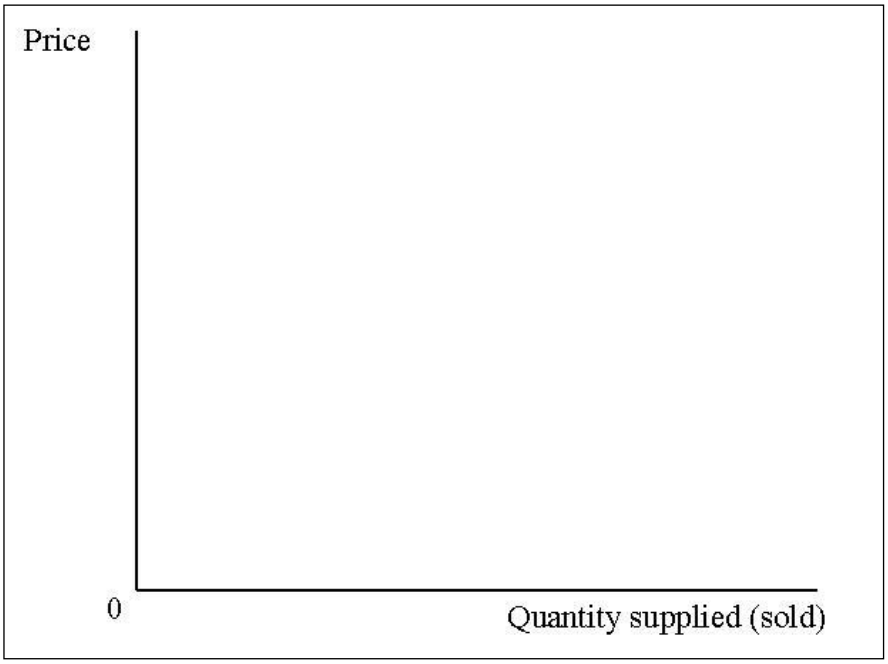


Figure 1.3a. A blank graph for you to show the relationship between price and the quantity supplied for a normal good or service.

Hopefully, if you have followed the previous discussion, your graph should look something like this:

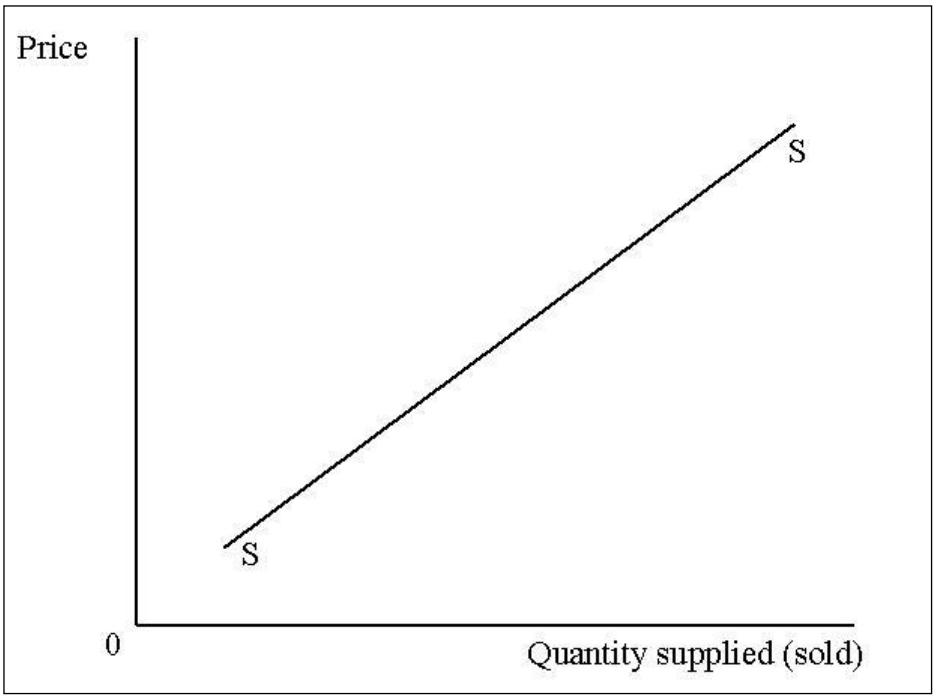


Figure 1.3b. A graph to show the relationship between supply and demand in a market situation.

As with the first graph referring to demand, I would hope that this second graph, which refers to supply, is pretty much common sense again. As the price of a good or service increases, then more of that good or service will be produced. What is important is the issue of why. As the price increases then firms that before could not cover their costs at the lower prices can now do so, hence they enter the market and the quantity supplied increases. If we stick rigidly to our assumption of profit maximisation, then more accurately as the price rises, potential profits in this industry also rise, and hence firms from other industries may switch production to this good or service because this now offers the best opportunity to maximise profits. A second factor is that as the price rises existing firms will also produce more, as this again is consistent with the assumption profit maximisation. The actual price of the good or service is therefore the first determinant of supply and shown by the supply curve. In addition the price however, there are six other basic determinants of supply. These are:

- Costs of production
- Government Policy (taxes and subsidies)
- The price of other goods that can be made with the same factors of production
- The price of goods in joint supply
- Natural shocks
- Aims of the producer

The cost of production will affect the level of supply. As we have seen already, if the price is relatively low, then there may exist few suppliers who can actually make a profit out of producing the good. If follows therefore, that if the costs of production were to rise then this would cause a decrease in supply (at all prices), as, other things being equal, there is less

profit being made on the good or service. Similarly, a reduction in the costs of production will cause an increase in supply at all prices for the opposite reasons. The second factor listed above is Government Policy, and in this case this mainly relates to the government's use of taxes and subsidies. As noted above, virtually all economies in the world today are of the mixed economy variety. One of the major ways in which the Government influences the market is through its use of taxes and subsidies. In essence, this has the same effect as the cost of production on the quantity supplied. The imposition (or increase in the level) of a tax will effectively increase the costs of production and cause a fall in supply at all prices. The payment (or increase in the level of) a subsidy to a service provider has the net effect of reducing costs, and hence we would expect an increase in supply at each and every price. The obvious example is the subsidy paid to the passenger rail operators. In 2006/7, this amounted to a total of £1.1bn[1]. The net effect of payment of this subsidy was to reduce the net costs of production and hence enable these companies to make a profit through providing rail services. As a final point, note that whilst here we have illustrated the effect of taxes/subsidies on the level of supply, other forms of government intervention exist, mainly regulation, that will impact upon the level supplied to the market.

The fourth factor listed above is the price of other goods and services that can be made using the same factors of production. Given that our producers are assumed to profit maximise, then if the price of any good or service that could be produced using the same factors of production was to rise, our producers are likely to switch production to that particular market. This would cause a reduction in the level of supply at each and every price for the current good or service. As we already know, all resources are scarce (limited), hence in order to enter the new market producers must find the resources from somewhere else (i.e. some other market). A rather contrived example relates to scheduled and charter air services. These are viewed as quite different markets, however if the price of scheduled services was to increase, then some operators in the charter market may decide to switch production to scheduled services. The rise in the price of scheduled flights would therefore cause a decrease in the supply of chartered flights.

The fifth factor given above is the price of goods in joint supply. In simple English, this means the price of goods that are produced at the same time. The best known example is the refinement of crude oil. To simplify what is undoubtedly a far more complicated process, at different temperatures different forms of 'fuel' are refined off. At around 200 degrees centigrade for example, lubricating oil is distilled off. At 40 degrees however, petroleum is distilled off. Thus if there is an increase in the price of lubricating oil, we may expect an increase in the supply of petroleum at each and every price as this is an automatic by-product of producing lubricating oil. You may think that producers would simply throw away unwanted by-products (although probably not petrol!), but as in theory our firms are profit-maximisers, they would attempt to sell any by-product for which there is a market.

Natural shocks simply relates to natural disasters such as the weather, flood, drought, pest etc, or abnormal circumstances arising from war, fire, political events and so on. The obvious example is the effect of a drought on the production of arable produce.

[1] Figure taken from TAS Rail Industry Monitor 2008, and note it excludes £3.5bn paid directly to Network Rail in respect of passenger services..

The final factor listed above is the aims of the producers. We have already stated and underlined that the aim of the producer is to maximise profits. This however may be a long-term aim, or may be an aim that may be pursued in a number of different ways. If there was a switch in the emphasis of the aims of producers, then we may expect a change in the level of supply in the market. For example, a firm may decide that in order to maximise profits in the long run it needs to expand its market share in the present. This would be known as sales maximisation, which may lead to quite different behaviour from firms and any change to a sales maximisation goal would almost certainly change the conditions of supply. In this case, we would expect an increase in the level of supply at all prices as companies are now willing to accept lower profitability levels in the present to expand market share in the future. Another example is that of railway companies, who are said to pursue a policy of sales maximisation. This is not for the reasons given above however, but because such an aim is consistent with profit maximisation. In very simple terms, the cost of carrying an extra passenger on a train to the train company is extremely small. Any revenue therefore that can be obtained from that extra passenger will virtually all be profit. The aim of the train company therefore is to attempt to fill all of the available capacity, or in other words, sales maximise.

Another aim that may be pursued by a company is to revenue maximise. That is, extract as much revenue from the market as possible. This is not the same as sales maximise, because this may be done in some cases by restricting supply, hence the good becomes more scare, and consequently this puts upward pressure on prices, causing them to rise. We will return to this issue when we examine the issues of demand elasticity and price discrimination.

As with demand, a change in any of the above factors are known as a change in the conditions of supply, and are shown on our supply curve diagram by a shift in the supply curve. In simple terms, any change in any of these factors will change the underlying relationship between price and the quantity supplied. This is illustrated in figure 1.4 below, again using the market for rail services. In this example, this illustrates what would happen if there was a reduction in the level of subsidy paid to operators.

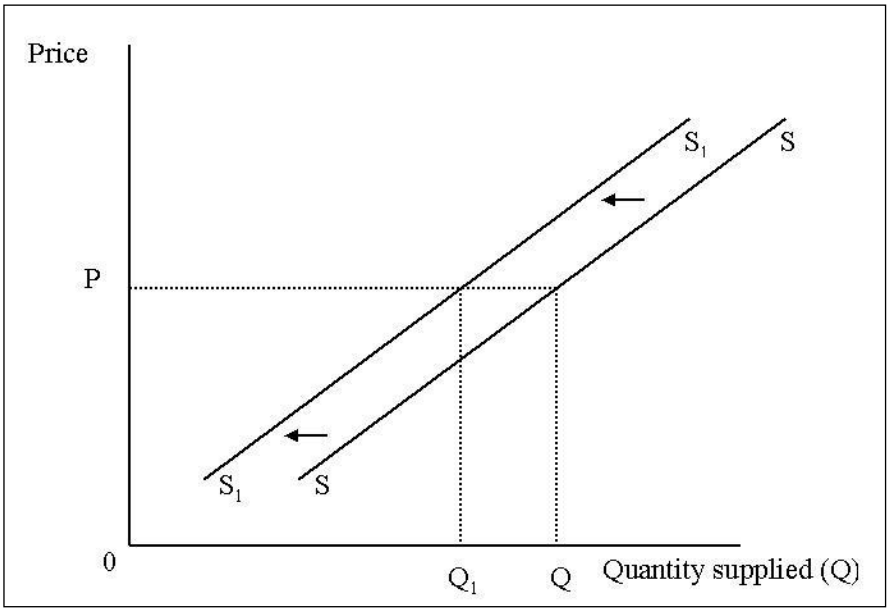


Figure 1.4: Change in the conditions of supply for rail services shown by a shift in the supply curve (a decrease in supply)

In figure 1.4, at the original price of P the quantity supplied for rail services is given by Q. With a reduction in the level of subsidy paid to rail operators, then other things remaining equal, we would expect less rail services to be supplied at each and every price. Thus at the price P, the quantity supplied would fall to Q_1.

Although we have used a change in the level of subsidy paid to rail operators to illustrate shifts in the supply curve, the same logic applies to each of other five determinants of the conditions of supply introduced above. You should now attempt Self Assessment Exercise 1.2 at the end of this unit.

One final note on demand and supply analysis before we go on to look at the market. You will have noted that in several parts the term 'other things remaining equal' has been used. This applies to all of our determinants of demand and supply, and is sometimes given in its Latin equivalent of 'ceteris paribus', although such terminology sounds a bit antiquated and pompous these days. What it means is exactly what it says, all other things remaining equal. Thus our analysis of the effect of an increase in income for the demand for a good, is only true if all other things remain the same. This is one of many underlying assumptions associated with Economics, but it does allow us to isolate the issue being examined and come to some logical understanding of the topic. If all other things did not remain the same, then the list of possible outcomes would be endless and we would have to spend our lifetimes studying economics. Fortunately, most of us have better things to do with our lives!

THE MARKET

Having illustrated the concepts of supply and demand, we are now in a position to put the two concepts together to look at 'the market'. As you will have noticed, both supply and demand curves were illustrated with the use diagrams that had the same labels on the vertical and horizontal axis. Namely, price was shown on the vertical and quantity on the horizontal. As a result, we can show these two concepts on the same graph, which is done in figure 1.5:

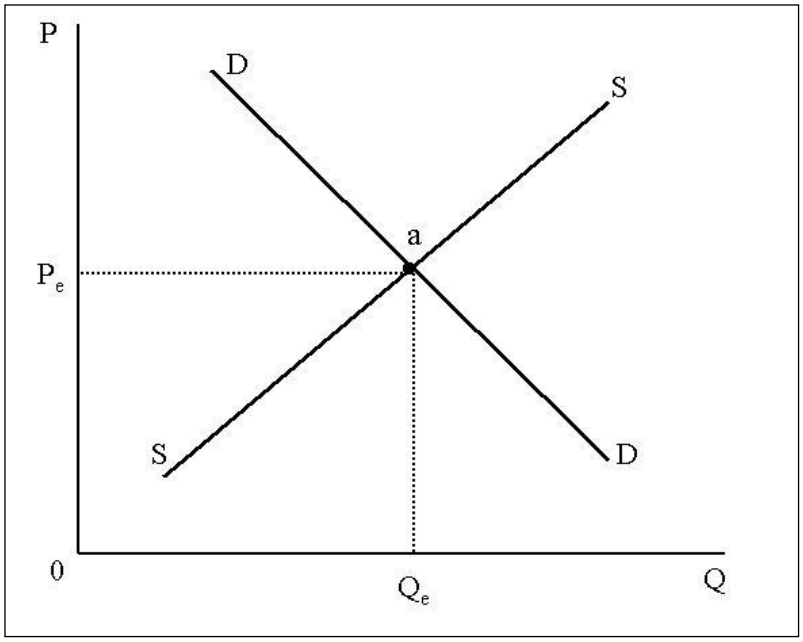


Figure 1.5: A graph to show the relationship between supply and demand in a market situation.

As before, demand is shown by the line labelled D and supply by the line labelled S. These two lines intersect at point a and at this intersection we have a market price of P_e and a quantity of Q_e. This is known as the equilibrium price and quantity, i.e. the point at which the market is in balance. Is it as simple as this? In a certain sense it is, but we need to know why. For example, one possibility for our two lines intercepting may be purely because we drew them that way. Consider the following example:

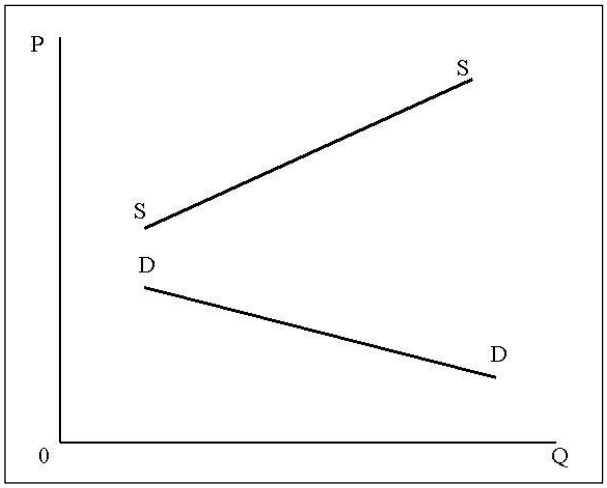


Figure 1.6: A graph to show that demand and supply need not intersect.

We have done nothing clever here, simply drawn our demand and supply curves in different positions. In this case however, they do not intercept. If we stick with economic analysis and thus believe that these lines actually mean something rather than simply being meaningless drawings on a piece of paper, then we need to explain this diagram. In this example, the diagram shows that the highest price that consumers would be willing to pay for this good or service is below the lowest price that suppliers would supply the market with this good or service. There is therefore no 'meeting' (through the price mechanism) of consumers and producers and thus there is no market for this good or service.

Let us return to our previous diagram, and this time think about what we have illustrated (and it is not simply meaningless lines drawn on a page of paper!).

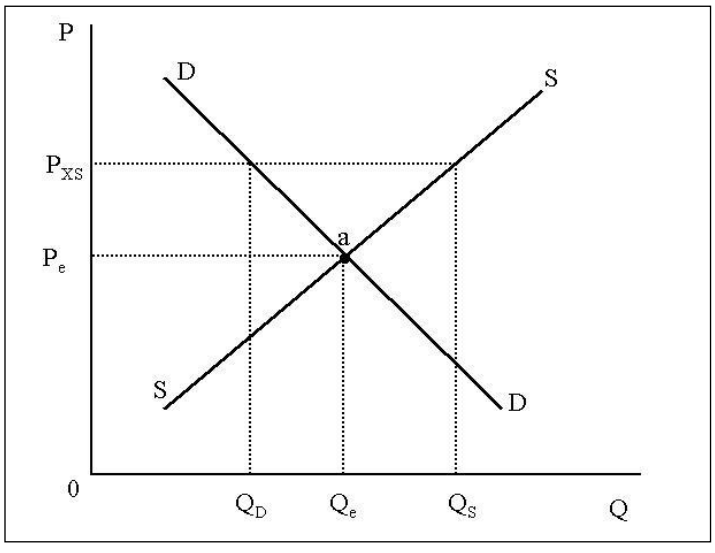


Figure 1.7: Excess supply in the market.

We have drawn our diagram of the market as before, except this time in order to help focus our thoughts we have added another price, that at level P_{XS}. At this price level, the quantity that would be supplied would be Q_S, whilst the quantity that would be demanded at that price would be Q_D. As you can see, the quantity supplied by far exceeds the quantity being demanded. This is known as excess supply, and because not all that is being produced is being consumed, there is either a high degree of wastage or stocks pile up. Using the first case, if this was say bus services, this means that there would be a very high level of capacity that the producer is not receiving a financial return for i.e. empty seats. In some cases this will result in loses, either of the entire bus company or on particular routes or services. Several things are likely to happen therefore – some companies are likely to go out of business, hence reducing the quantity supplied, whilst other companies are likely to withdraw unprofitable services, hence again reducing the quantity supplied. The level of excess supply would therefore be reduced. Alternatively (or additionally), suppliers may reduce the price in an attempt to fill the excess capacity, hence the quantity demanded would increase. These processes would continue until the market was in balance, which only takes place at price P_e. At price P_e, the market is said to be in equilibrium (balance). Consider figure 1.8.

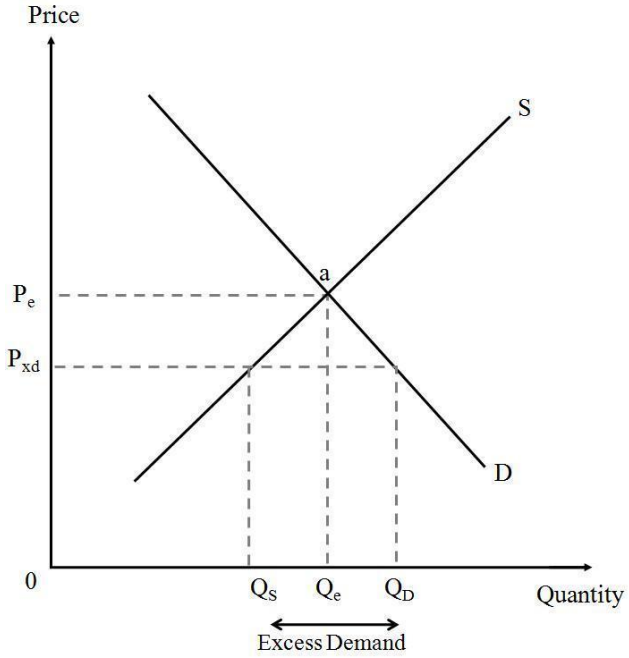


Figure 1.8: Excess demand in the market.

Figure 1.8 shows our market as before, except this time we have added a price level, shown by P_{XD}, at below the market equilibrium price. In this example, the quantity demanded, Q_D, far exceeds the quantity supplied, Q_S, and we have what is known as excess demand, shown by Q_D - Q_S. In this case there exists a consumer demand that is not being met by producers. Using our example of bus services again, this would be most vividly illustrated by over-crowding on buses and long cues at bus stops. In order to meet such measures, some suppliers will increase the quantity being supplied. This however can only be done if the price increases. These 'extra' buses (at least in the short term), will have to be found from other sources, namely, other routes. As our producers are profit maximisers, then this will only take place if a higher level of return can be earned from switching the buses to this

route. In this case there is a sufficient demand for such a switch, hence this would occur. The increase in price however will put some of our consumers off, as being utility maximisers, the use of the bus may no longer represent the best value of their (limited) income, hence they may either not travel at all or use other forms of transport. Either way, the quantity demanded falls. As illustrated before, this process will continue until all excess demand is burnt off. This only takes place at price P_e.

Hopefully now, we have a basic understanding of the workings of the market through the price mechanism and the concepts of excess supply and excess demand and why there is an in-built tendency for the market to be in equilibrium. This was illustrated through the use of diagrams, but again hopefully these have provided clarity of thinking rather than confusion and disillusionment!

THE WORKINGS OF THE MARKET

We are now ready to finish this unit by putting all of these ideas together to analysis the impact upon the market price and quantity if there is a change in any of the determinants of supply and demand. We will continue to use the example of the market for bus services, as shown by figure 1.9

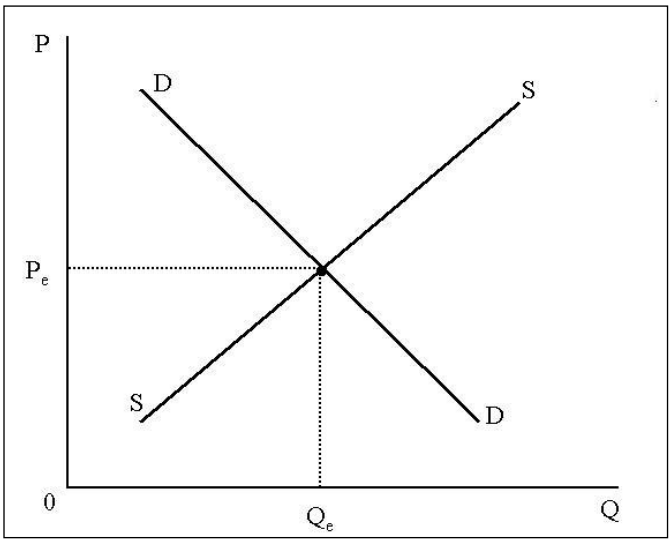


Figure 1.9: Illustration of the market for bus services.

This is our basic illustration of what should now be familiar to you of the basic market. In this case however we are using it to illustrate the impact of various changes in the conditions of demand and supply on the price and quantity traded in the market for bus services. Firstly, let us consider a demand side factor, namely, the effect on the bus market of a change in the price of a substitute service, a rise in the price of rail services. This would affect the demand side of the market as illustrated by a shift in the demand curve for bus services to the right i.e. an increase in the quantity demanded at each and every price. This is shown in figure 1.10.

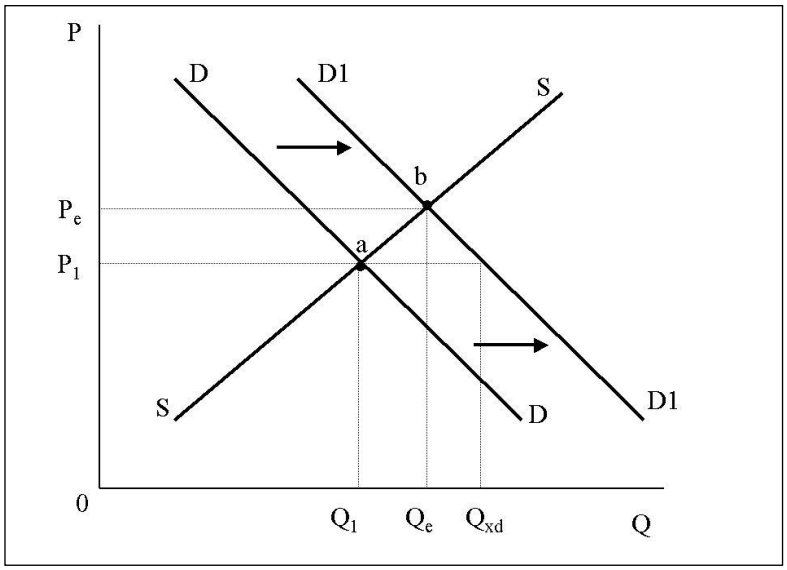


Figure 1.10: Illustration of the effect of an increase in the price of rail services on the market for bus services.

We begin with the market for bus services being in equilibrium (demand = supply) at point a with price P_1 with a quantity traded of Q_1. The price of a substitute, rail travel, then increases. As rail travel then represents less value for money or becomes out with the budget of some travellers, some switch to using the bus. This change in the conditions of demand for bus services is illustrated in figure 1.10 by a shift of the demand curve to the right i.e. an increase at each and every price, shown by D1. At the existing equilibrium price of P_1 therefore, there is now excess demand at that price. This situation has already been examined under figure 1.8. This excess demand ($Q_{XD} – Q_1$) is eradicated through firstly an increase in price by suppliers, which causes a reduction in the quantity demand (as shown by a movement along the new demand curve D1). As we have seen however, the increased price will also cause an increase in the quantity supplied (a movement along the existing supply curve S). This process continues until the market is back in equilibrium at a new market price of P_e with the associated quantity traded Q_e.

To illustrate the effect of a supply side factor, we will use the effect of an increase in the level of subsidy paid to bus operators. Such an increase would be represented by a shift of the supply curve to the right i.e. more would be supplied at all prices. This is shown in figure 1.11.

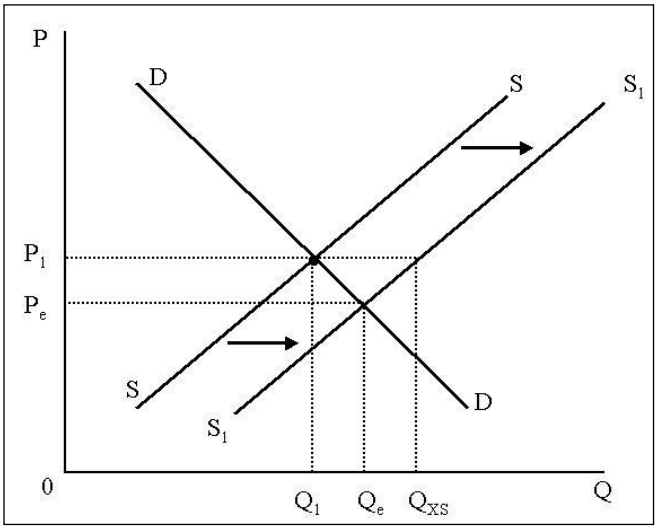


Figure 1.11: The effect on an increase in the level of subsidy paid to bus operators on the market price and quantity traded of bus services.

Prior to the increase in subsidy, the market was in equilibrium at price P_1 and quantity traded Q_1. The increase in subsidy is shown by a shift in the supply curve to the right (this is akin to a reduction in costs), hence at price P_1 there is now excess supply of Q_{XS} minus Q_1. This is the same situation that was illustrated with Figure 1.7. As before, the market would end up back in equilibrium at price P_e and quantity Q_e.

We could endlessly go on illustrating the effects on the various determinants of supply and demand on the market price and quantity traded, however these are left to the activities at the end of this unit. We should reflect however on what we have covered in this unit. Hopefully you have not felt too bombarded with diagrams, as you have seen we have illustrated the various concepts using eleven diagrams, quite a large number. In practice however, when we consider various factors on the market we would never contemplate as many as that, but only one at a time! At this stage however, you should become familiar with the workings of the market system through the price mechanism, and there is no substitute for practice, so let's do some!

Unit 1 Self-assessment exercises

Exercise 1.1

In the text you saw an illustration of the effect of a decrease in the price of bus services (a substitute good) on the demand for rail services. You should use the same process to illustrate what would happen in each of the following cases. In each example you should identify what type of determinant of demand is being examined and your reasoning for your answer (as there may be more than effect taking place).

- A rise in income on the effect of the market for rail services
- A fall in the price of a petrol in relation to the market for new cars
- The effect of issuing travel cards to students that entitles them to a 50% discount on buses on the market for bus services.
- The effect of an announcement that rail fares are to rise by 10% from the 1st of the month following the announcement on the market for rail services

Exercise 1.2

In the text you saw an illustration of the effect of a reduction in subsidy on the supply of rail services. You should use the same process to illustrate what would happen in each of the following cases. As above, you should identify what determinant of supply is being examined and your reasoning for your answer (as again there may be more than one effect taking place).

- A fall in the costs of production on the supply of bus services
- A change in the aims of rail operators from an aim of sales maximising to one of profit maximising on the supply of rail services
- A significant increase in the price of freight transport services on the market for passenger services
- The unlikely situation where the government decides to re-nationalise the entire rail industry on the supply of rail services.

Exercise 1.3

You should now consider each of the following factors on the price and quantity traded for the market highlighted. This time you should identify what side of the market (demand or supply) is being affected, what particular determinant has changed and you should explain your reasoning.

- The entry of a low cost airline into the long haul sector on the market for airline services.
- A rise in income on the market for bus services.
- An increase in fuel duty on the market for bus services
- A fall in the costs of production of bus services on the market for rail services

Exercise 1.4

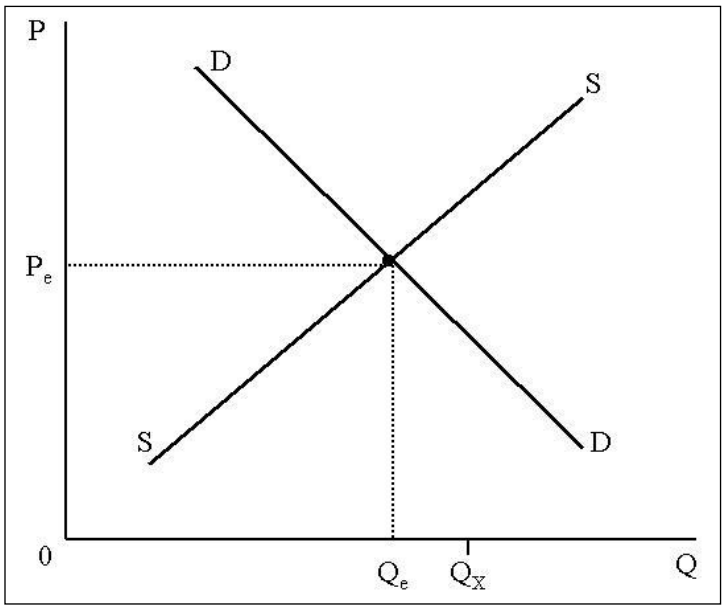
The full title of this module is 'Transport Economics and Appraisal". Using your knowledge of the economic concepts of scarcity, choice and opportunity cost, write a brief summary on what you believe are the links between Transport Economics and Appraisal Methods and what you believe the objectives (i.e. on what criteria should they be based) of any appraisal method should be.

Exercise 1.5

Re-examine the list that you drew up for the introductory exercise on key words that you believed to be associated with the study of Economics. Draw up a second sheet with three headings, Scarcity, Choice and Opportunity Cost. Now see if you can place each of your key words under one of these headings – are there any 'orphans' left??

Exercise 1.6 – Increasing the use of the railways

The Government recently stated as one of its aims for passenger rail services that it wished to see an increase in the use of rail travel. The simple question is, using your new found knowledge of economics, how can the government assist the market to achieve this aim. If we assume on the graph below that the government wished to see the quantity traded from the current position of Q_e increase to the level indicated by Q_X.



In the course of this exercise, you should consider both direct government intervention in this market, as well as indirect intervention that it may take in other transport markets. Let us further assume however that the aim would be to increase the level of rail travel without causing a modal shift from other public transport markets.. You should illustrate your answer with all relevant diagrams. (HINT – there are basically three possible scenarios, although you may come up with more!).

Of the scenarios that you have devised, which do you consider would be the most effective?

SUMMARY

In this unit, the first in the module on Transport Economics and Appraisal, you have had a general introduction into the study of Economics. Hopefully, you should see why this is an important area in the study of Transport issues. We have defined a number of common concepts within economics, which will be referred to later on in the module. The main determinants affecting the demand and supply of transport services were outlined and the basic workings of the market were illustrated. These elements will be built upon in the following units and from the self-assessment exercise you should now be able to understand the basic workings of the market.

FURTHER READING

The further reading for this unit is Cowie (2010) Chapters 1 and 3.

UNIT TWO
ELASTICITY OF DEMAND

INTRODUCTION

As we have seen in Unit 1, the demand for a good is "the number of units per period of time that consumers will purchase at any given price". Furthermore, we saw that the overall 'basket' of goods and services that consumers end up with is driven by their desire to maximise utility. That is that they seek to achieve the highest value for money from their limited incomes. We also saw that demand is linked to the supply of goods by transport operators to form a dynamic transport market. Demand is a result of consumers expressing their own preferences between goods at the relative prices that they face, taking into account all money available to them. This unit focuses on the factors behind the demand of individual consumers and specifically examines demand elasticity.

RECOMMENDED READING

The recommended reading for this unit is:

Cowie (2010), *Economics of Transport*, Chapter 4, Transport Demand Elasticity.

LEARNING OUTCOMES

Once you have completed this Unit, you will be able to:

* Explain the term 'elasticity'
* Outline in principle the main types of demand elasticity in transport markets
* Be able to calculate some basic elasticities and be able to distinguish between point and arc elasticities
* Link price elasticity of demand to operator revenue.

Elasticity of demand

As an introduction to the concept of elasticity, look at the return train fares on offer from Edinburgh to Leeds and Edinburgh to London as shown in Table 2.1.

Table 2.1. Train fare options from Edinburgh - departing on 15th December 2009 and returning on 17th December 2009[2]

Fare type (returns)	Fares Edinburgh – Leeds	Fares Edinburgh – London
Cheapest Single (X2)	£46.00	£108.00
Off-peak return	£76.60	£174.80
Anytime return	£172.00	£271.00
Cheapest 1st return	£269.00	£229.00
First dine and go	£269.00	n/a

There is a large variety in prices of (return) tickets from Edinburgh to Leeds (£46.00 to £269.00) and London (£174.80 to £229.00), for the same journey. We will examine the pricing policy of public transport operators in Unit 5, which will show why they charge different fare types for the same journey. However, the underlying reason for the difference relates to consumers, and in particular their different price elasticities of demand.

As a general rule, when the price of a good rises the quantity demanded will fall. In most cases we want to know more than this, we want to know by just *how much* the quantity demanded will fall. In other words, we will want to know how *responsive* demand is to changes in price, or in plain English, the level of price sensitivity. For example, how sensitive are people to purchasing train tickets if the fare was to rise by 5%, 10% or 50%? In other words, if you bought a train ticket for £100, would you still buy the ticket if the price was to rise to £105, £110 or £150? Price elasticity is therefore the responsiveness of passengers or potential passengers to the prices on offer.

However, elasticity can involve more than a response to price changes, it can concern other factors affecting demand. A general definition of the elasticity of demand is the "response of demand for a product to the change in one of its determinant factors".

[2] Source: *www.thetrainline.com*, accessed on 16th November 2009.

In the remainder of the unit we will outline the types of elasticity of demand, show how to calculate elasticity of demand, provide some examples of elasticity studies, and provide a graphical illustration of the elasticity of demand. Also, at the end of the unit we will see how the changes in price can be measured to determine the extra revenue that will be achieved by public transport operators through variation of prices.

Types of elasticity of demand

There are four main types of elasticity of demand that are relevant to the transport sector. Each will be described in turn with a simple definition and equation.

1. OWN PRICE ELASTICITY

The price elasticity of demand, as discussed previously, is the "responsiveness to changes in the transport operator's own price."

It can be calculated as:

$$\text{Price elasticity of demand} = \frac{\text{proportionate change in quantity demanded}}{\text{proportionate change in price}}$$

Or you may see this equation expressed in its short hand form of:

$$\text{PED} = \frac{\%\Delta D}{\%\Delta P}$$

Where:

PED	=	price elasticity of demand
%	=	percentage
Δ	=	change (represented by the Greek letter 'delta')
D	=	quantity demanded
P	=	price

Price elasticity therefore is an assessment of the relative changes in the quantity demanded and price, and as such, provides an indicator of the price sensitivity of consumers[3]. To give a hypothetical example, if an operator were to increase its fares by 40% but the quantity demanded was to fall by only 6%, then price elasticity would be -0.15, as calculated by:

[3] Notice that in all of these examples we are using a comparison of percentage changes to calculate the rate of change of demand to the rate of change to price. For this reason, these are known as 'arc' elasticities, as the elasticity value is calculated over a range of the demand curve. Alternatively, the rate of change can be found at any particular point by taking the differential, and this would be known as 'point' elasticity.

$$-0.15 = \frac{0.06 \text{ (proportionate change in quantity demanded)}}{0.4 \text{ (proportionate change in price)}}$$

Note that the sign of the price elasticity value in this case is negative. A *rise* in price (a positive figure) usually causes a *fall* in the quantity demanded (a negative figure). Similarly a *fall* in price usually causes a *rise* in the quantity demanded. This is our basic law of demand from Unit 1. Thus when working out price elasticity of demand, we either divide a negative figure by a positive figure, or a positive figure by a negative figure. Either way, we usually end up with a negative figure. This of course applies to our definition of a 'normal' good, where a rise in the price will cause a fall in the quantity demanded. For a normal good therefore, as in this example, the sign of price elasticity will always be negative[4].

What does this figure mean however? Own price elasticity is a formal measure of the rate of change of the quantity demanded in comparison to the rate of change of price. Where the price elasticity is greater than one therefore, then this means that the proportionate change in demand is greater than the proportionate change in price. Where it is less than one however, then the proportionate change in quantity demanded is less than the proportionate change in price. To use our price sensitivity terminology, then where the price elasticity is greater than one then consumers are relativity price sensitive, however where it is less than one then consumers have a relatively low level of price sensitivity.

This important division at the unitary value (i.e. one) highlights an important division within the concept of elasticity of demand. *Elastic demand* is where the quantity demanded changes by a larger percentage than price i.e. consumers are relatively price sensitive. The most extreme case would be if the price was to change up or down by any degree, demand would fall to zero. This is illustrated in figure 2.1:

[4] Contast this with what is known as a 'Giffin' good, which exists where an increase in the price of a good or service actually brings about an increase in the quantity demanded, not a decrease. This may apply for example to goods such as cheap jewellery, where a price increase may lead to a perception of higher quality, and hence quantity demanded increases. In such cases, the price elasticity of demand would be positive. Giffin goods however tend to be very rare.

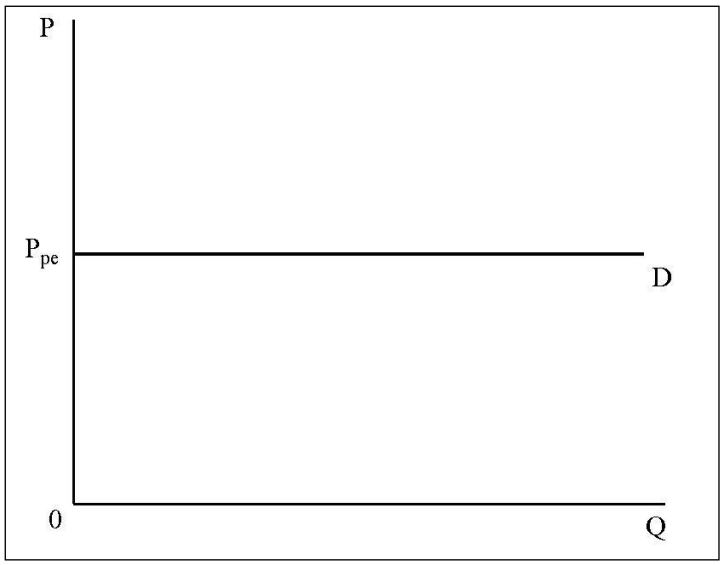


Figure 2.1: Perfectly price elastic demand

This is known as perfectly elastic demand, and whilst in this context is purely a theoretical concept, it does have practical implications that we will examine later in the course. Importantly however, it gives us a top barrier for our range of price elasticity. If for example, this situation had been brought about by a 15% increase in price, then price elasticity of demand would be given by:

$$\text{PED} = \frac{\%\Delta D}{\%\Delta P} = \frac{-\infty}{15\%} = -\infty$$

Therefore the top end of the range is negative infinity (as given by the symbol ∞), but not beyond!

The other extreme case is where the change in price will produce absolutely no change in the quantity demanded i.e. consumers have zero price sensitivity, as they will purchase exactly the same quantity of the good whether the price rises or decreases. This extreme case is known as perfectly price inelastic, and is shown in figure 2.2.

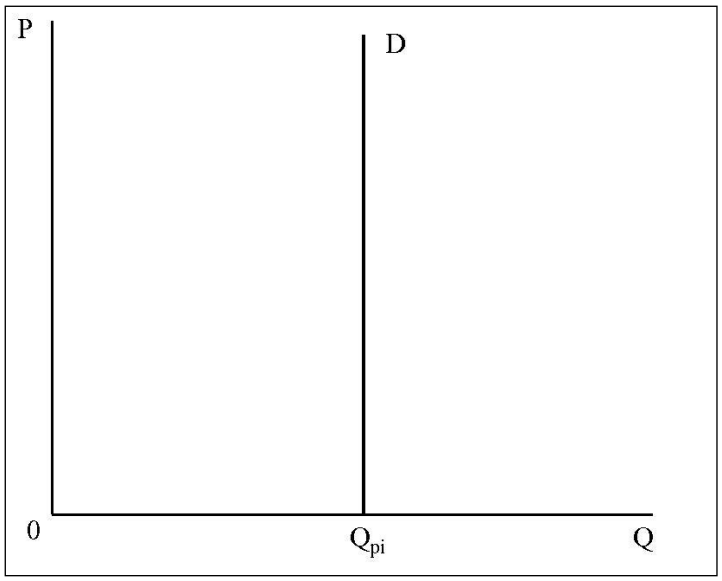


Figure 2.2: Perfectly price inelastic demand

Where demand is perfectly inelastic, then no matter the price consumers will purchase exactly the same quantity of the good or service i.e. the demand curve is completely vertical. Again this has few practical applications, although the demand for salt will certainly be highly inelastic. Using our 15% example of an increase in price, then the equation for calculating the value of perfectly inelastic demand would give us:

$$\text{PED} = \frac{\%\Delta D}{\%\Delta P} = \frac{0\%}{15\%} = 0$$

This therefore gives us a bottom value of zero. A final example is where the relative change in price is exactly matched by the relative change in quantity demanded, this is known as unitary elastic demand, and is shown in figure 2.3.

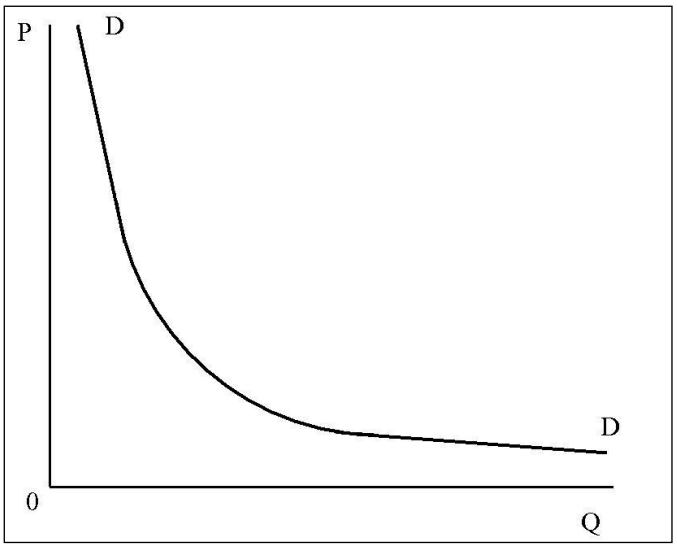


Figure 2.3: Unitary Price Elastic Demand

Note in this diagram we have relaxed our quick method of drawing demand curves as straight lines as we cannot shown unitary demand as a straight line as the level of price elasticity will vary along the different points on any straight line.

To quickly summarise then, the relatively level of price sensitivity of consumers is measured in economics through the concept of elasticity. It is important because the basic law of demand tells us that if the price rises then quantity demanded falls and vice-versa. What is of real use however is the extent to which the quantity demanded falls or rises in relation to changes in price, and this is what price elasticity assesses. We saw that this can be broken down into two/three different classes of price elasticity, and these are summarised over the page in figure 2.4, which actually expands this initially three types into five.

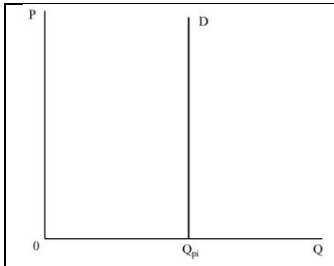
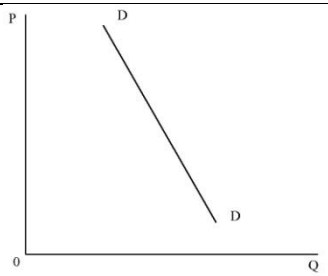
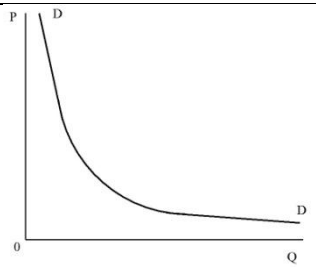
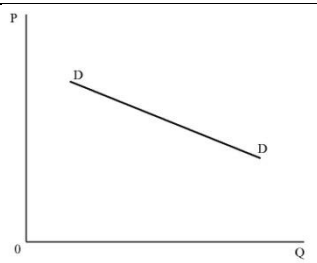
				
Perfectly Inelastic Demand	Relatively Inelastic Demand	Unitary Elastic Demand	Relatively Elastic Demand	Perfectly Elastic Demand
PED: 0	PED: 0 to -1	PED: -1	PED: -1 to -Infinity:	PED: -Infinity

Figure 2.4: Summary of Price Elasticity of Demand

We have therefore seen that price elasticity of demand is a mechanism for assessing the extent to which consumers will react to changes in price. Has this got any other practical implications however? The answer of course is yes, but first requires a simple lesson in algebra before we examine it.

Put in simple terms, a company's total revenue from selling a good or service can be found by simply multiplying the quantity sold by the price at which each unit was sold. Therefore, if we were to sell 100 units at £5 each, then the total revenue would be £500. In a more generic form, this could be written as P * Q, where P is the price and Q is the quantity sold. This can be illustrated on our demand curves in the following way:

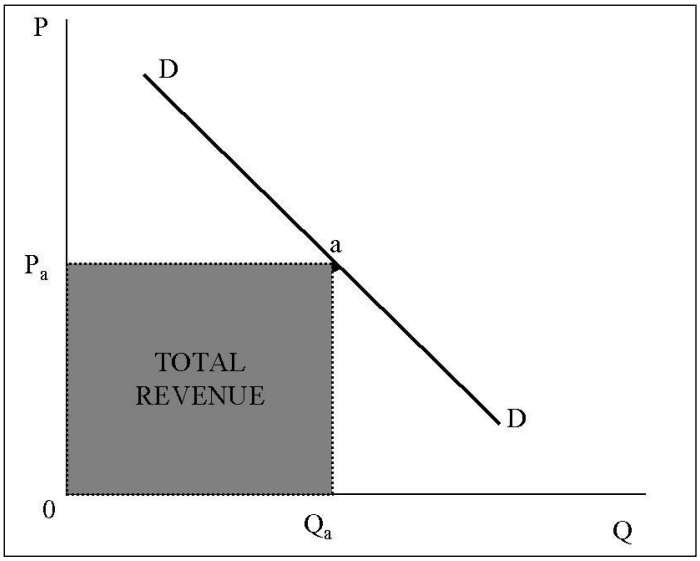


Figure 2.5: Illustrating total revenue using demand curves

Basically, the area of the rectangle outlined by points 0, P_a, a and Q_a and shown in light grey in figure 2.5 is the total revenue received from selling at price P_a. If for example the price was £5 and the quantity sold was 100 units, then the area, given by length (100) times breadth (5), would be £500. If the price was to increase to P_b, then we could show the gain and loss in revenue such a price change would bring about. If the firm was to increase price, it would sell less units (basic law of demand), but it would receive more per unit sold. This is illustrated in figure 2.6 below.

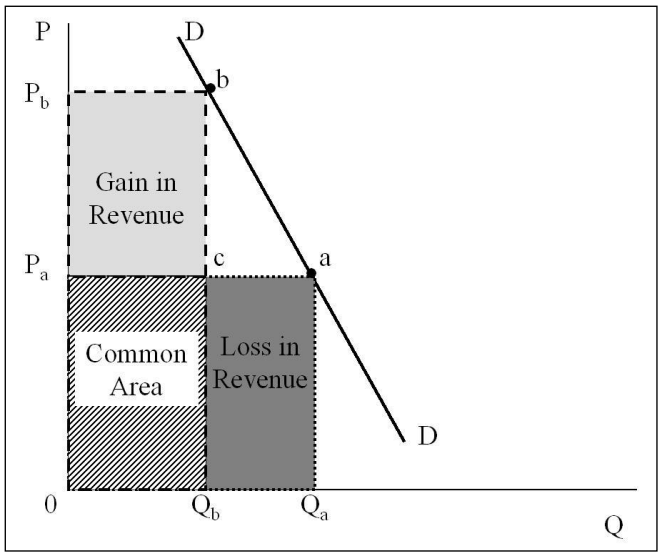


Figure 2.6: Illustrating changes in total revenue using demand curves

This diagram looks considerably more complicated than the previous one, but let us examine it in a bit more detail. Firstly at price P_a (which we examined above), the company sells quantity Q_a, and hence the rectangle outlined by the area 0, P_a, a, Q_a was shown to represent the total revenue to the firm. Using the same logic, if the firm was to sell at price P_b, then it would sell quantity Q_b and the total revenue would be given by the rectangle outlined by 0, P_b, b, Q_b. What you should be able to see is that there is a common area shared by these two different scenarios. That is the area that is cross-hatched above and labelled the 'Common Area'. In effect, that proportion of revenue will accrue to the firm if it applies price P_a or price P_b. If however the firm was to increase its price from P_a to P_b, then it would not receive in revenue the area labelled 'Loss in Revenue' as it would be selling less units overall. For this loss however, there would be again on the extra revenue received for each unit it did sell, and this is labelled 'Gain in Revenue' in figure 2.5. What the key issue surrounds is if the gain in revenue is greater than the loss. If this is the case, then overall revenue will increase. You should be able to see that in this example the answer would be yes, as the area of the gain is smaller than the area of the loss. Hence increasing price from P_a to P_b will lead to an increase in total revenue. This is because demand is relatively inelastic.

Note that elasticity of demand is about changes in price and the effect such changes have on revenue. At this stage, we have no idea how this may affect profitability, as we have as yet not examined the supply side of the market. For example, if in the previous example the price had been cut from P_b to P_a, then this would have decreased total revenue. However, we do not know what may happen to unit costs from increasing supply from Q_b to Q_a, thus we cannot say anything about profitability. If unit costs were for example to reduce significantly, then such a price cut although reducing total revenue may actually increase total profit.

To further illustrate, consider the following example where demand is relatively elastic. This is shown in Figure 2.7.

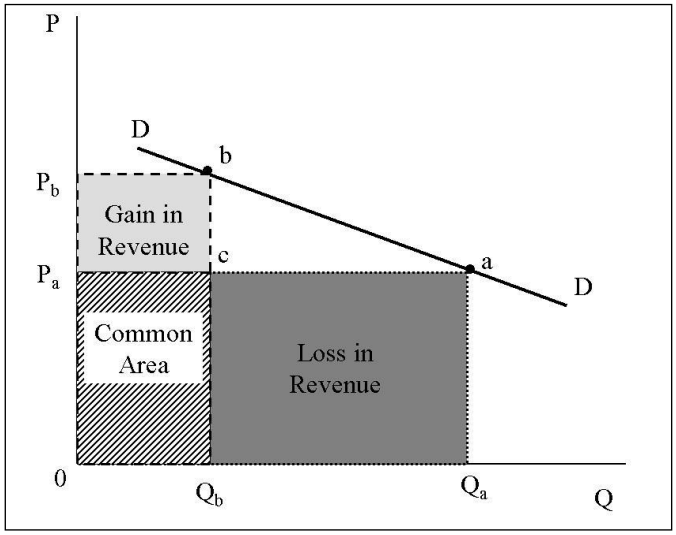


Figure 2.7: The effect on revenue of price changes of a relatively elastic good.

In this example, a rise in price from P_a to P_b will reduce demand from Q_a to Q_b, with the loss in revenue given by the area Q_b, Q_a, a, c and the gain shown by the area outlined by P_a, c, b, P_b. In this case, we see that the increase in price has led to a decrease in total revenue. Note again that if the price change had been the other way around, then total revenue would have increased.

To summarise therefore, where demand is relatively inelastic then price increases will increase total revenue, whilst price cuts will decrease total revenue. Where demand is relatively elastic, then price increases will lead to decreases in total revenue whilst price cuts will lead to increases in total revenue.

DETERMINANTS OF PRICE ELASTICITY OF DEMAND

There are three basic determinants of price elasticity of demand. Before we examine them however, confusion can arise between determinants of price elasticity of demand and the actual determinants of demand itself, such as income, the price of substitutes and complementary goods that we examined in unit 1. The important distinction is that the determinants of demand are the factors that determine the quantity of goods or services that consumers will purchase at a given price (and as such represent the conditions of demand). The determinants of price elasticity on the other hand are those factors that determine the extent to which consumers will change the quantity demanded should the price change, or alternatively are factors that affect the degree to which consumers are price sensitive or not. .

To return to the determinants of price elasticity of demand, as highlighted above there are three basic determinants, these are:

- The number and closeness of available substitutes
- The proportion of disposable income used to purchase the good
- The time dimension

The higher the number of substitutes and the closer they are in meeting the same need or want, the higher will be the price elasticity for a good or service. This should be fairly obvious. If I use the Blue Bus Company's service to travel to work, and Blue Bus should increase their price, then I am far more likely to switch to an alternative if that alternative is readily available. If there is an alternative, say the Red Bus Company's service left from the same stop one minute later, then this would be almost a perfect substitute for Blue's service, and hence I would make the switch. If however Blue's service was the only one available to me, then I would have no alternative but to pay the higher price (or quit my job!). This for example is one reason why the price of petrol varies very little from petrol station to petrol station, because there are a high number of available substitutes (Esso, BP, Shell, Jet, Safeway, Sainsbury etc etc), and it is all basically the same stuff that they are selling. Hence there is a very high level of substitutability, and if one company was to increase its price, then consumers can easily switch to one of the other providers.

The higher the proportion of disposable income[5] spent on the good, then the higher the price elasticity of demand. If consumers' are about to part with a large amount of their income on a good or service, then they are likely to shop around first to ensure that they are getting the best deal. No one however, would shop around before purchasing a chocolate bar because it simply is not worth it – the amount saved is likely to be marginal and certainly less than the cost involved (i.e. time) of ensuring that the lowest price chocolate bar has been purchased.

The final determinant of price elasticity is the time factor. In the short term, in many instances individuals are tied in to using a certain company's products or services, however time may bring about a change in behaviour or other factors. Using our Blue Bus Company example again, while this was the only service in the short term I would have to use their service, however in the longer term I may decide to purchase a car or alternatively the Red Bus Company may decide to start up a competing service. Time therefore is an important dimension in determining price elasticity of demand. Also related to time is if the journey is essential or not. An essential service, such as where commuters have to travel into the centre of a city, will have a relatively inelastic demand as opposed to non-essential journeys, such a family day-out to the countryside, where there is a far higher degree of flexibility.

[5] Disposable income is net income, i.e. income after tax has been paid.

Much research has been carried out in the transport industries to determine the level of price elasticity. In the short run elasticity of demand is likely to be relatively inelastic (-0.1 to -0.4) because the locations of journey origins and destinations are fixed and consumers have no alternative but to make the journey. In the longer term other factors determining elasticity can vary. People can change jobs or housing, buy a car, or other operators can appear on the scene.

Elasticities are generally found to increase the longer the journey under consideration[6]. This is not simply a function of distance, but rather a reflection of the absolute magnitude of, say a 10% rise on a £5 fare compared with that on a £500 fare (i.e. the proportion of disposable income spent on transport). Longer journeys are made less frequently, and thus people gather information about prices in a different way. They also tend to involve leisure rather than business travel.

Goodwin (1992) undertook an extensive compilation and assessment of 50 elasticity studies for car and bus travel in Europe from 1980 to 1990[7], and despite its age this still remains the key study in this area. For urban bus travel, the review calculated an average value of price elasticity (impact of fares on patronage) of –0.41, but indicated a wide variation between the short-term and long-term impacts. As –0.41 falls in the range 0 to –1, this would be classified as being relatively inelastic demand. Thus a 1% rise in price would bring a smaller proportionate change in the quantity demanded, in this case a 0.41% fall.

An interesting point to make is that the price elasticity of demand will depend largely on the perception that consumers have of the price of a journey, which may differ from the actual monies expended. Users of different forms of transport (or sometimes different services of the same mode) are often confronted with different methods of payment, which makes price comparisons difficult.

Price elasticity of demand is not the only type of demand elasticity however, and we now turn to examine the other types.

2. CROSS PRICE ELASTICITY

We have examined the (own) price elasticity of demand in some depth, and all of the concepts that we have examined apply to the other types of elasticity of demand. Rather than the sensitivity of consumers to changes in price however, it is the sensitivity of consumer demand in relation to a change in one of the other determinants of demand, whether that be income, price of other goods, advertising etc. We examine two of these. The first is known as the cross price elasticity of demand and this is the "measure of the

[6] Button K J (1993), *Transport Economics*, p44.

[7] Goodwin P (1992), *A review of new demand elasticities*, Journal of Transport Economics and Policy 26, pp 155-170.

effect of a change in the fares or rates of one operator on the demand for the services of another".

It can be formally calculated as:

$$\text{Cross elasticity} = \frac{\text{percentage change in quantity demanded of our service}}{\text{percentage change in price of another service}}$$

Or using a shortened formula:

$$\text{CPED} = \frac{\%\Delta D_A}{\%\Delta P_B}$$

Where:

CPED	=	cross price elasticity of demand
%	=	percentage
Δ	=	change (represented by the Greek letter 'delta')
D_A	=	quantity demanded of good A
P_B	=	price of good B

As can be seen from the formula, examination of cross price elasticity of demand involves examining two goods or services. Within the transport sector, these two services could be examined at different levels. This could be the cross price elasticity between two different transport modes, such as the train versus the car. Secondly, cross price elasticity could be calculated within the same mode, such as East Coast Mainline Glasgow to London rail service versus Virgin's West Coast Mainline Glasgow to London rail service. Finally, it could be examined within a single operator if they offer a variety of fares for the same journey but different standards of service, such as the first class fare/service versus the standard fare/service.

The cross price elasticity of demand allows us to distinguish between substitute goods and services and complementary goods and services. If the effect of a price increase in one good has a positive effect in terms of the demand for another good or service, then these would be considered to be substitutes. Say for example a reduction in the subsidy paid to rail operators caused an increase in the price of rail services, then we may expect the following:

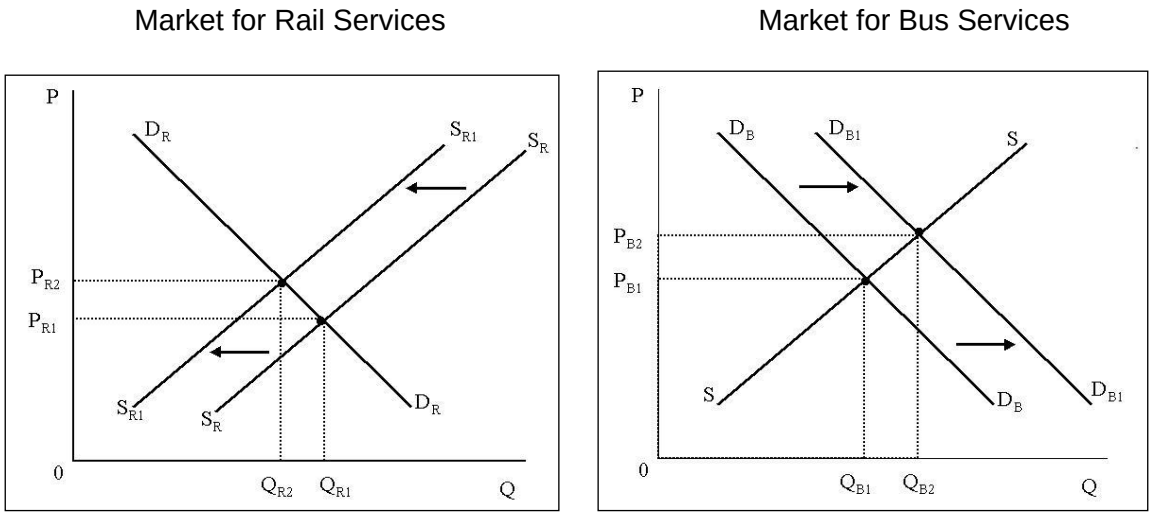


Figure 2.8: Cross Price Elasticity of Demand, substitutes

This is our standard demand and supply curves, with the reduction in subsidy to rail operators effectively representing an increase in costs, hence the supply curve shifts to the left and eventually results in an increase in the price of rail services. As rail is a substitute for bus travel, then this increase in the price of rail causes an increase in the demand for bus services at each and every price, as shown on the right by a shift in the demand curve. You should be able to clearly see that P_{R2} is greater than P_{R1}, thus when the difference is expressed as a percentage, because P_{R2} minus P_{R1} is positive, this will be positive. Similarly, Q_{B2} is greater than Q_{B1}, thus again when expressed as a percentage, because Q_{B2} minus Q_{B1} is positive, this will be positive. When we roughly put these values into our equation for cross price elasticity of demand (between bus and rail travel), then we obtain:

$$\text{CPED} = \frac{\%\Delta D_B}{\%\Delta P_R} = \frac{^{+}ve}{^{+}ve} = {}^{+}ve$$

Where:

$\%\Delta D_B$	=	percentage change in quantity demand of bus services
$\%\Delta P_R$	=	percentage change in the price of rail services
^{+}ve	=	a positive value

By the same logic, the cross price elasticity of demand will be negative for goods and services that are complements. Consider the following almost classic example of the price of cars and the market for petrol. If the cost of manufacturing motor cars was to rise, then this would cause an increase in the price of cars and a reduction in the level of

quantity demanded. If there are less cars on the road, then less petrol will be required, hence there will be a decrease in demand for petrol. These are illustrated in Figure 2.9.

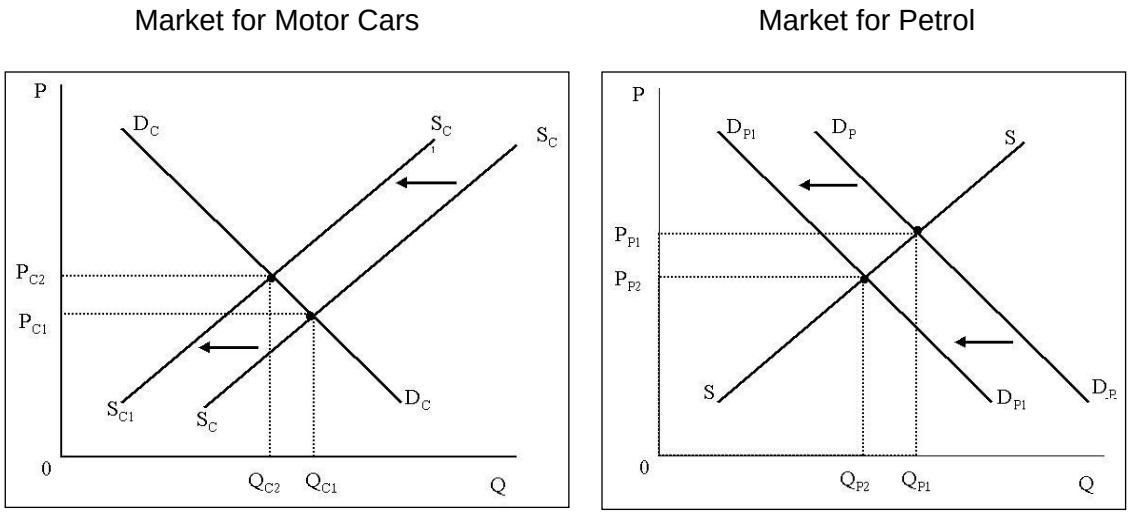


Figure 2.9: Cross Price Elasticity of Demand, complements

Again, this is illustrated by our standard demand and supply curves, with the increase in the cost of producing motor cars shown by the shift to the right of the supply curve and resulting in an increase in the price. As cars and petrol are consumed at the same time, this increase in the price of cars changes the market conditions for petrol, causing a decrease in demand, as shown on the right diagram by a shift in the demand curve to the left. As with above, you should be able to clearly see that P_{C2} is greater than P_{C1}, thus when the difference is expressed as a percentage, because P_{C2} minus P_{C1} is positive, this will be positive. This time however, Q_{P2} is less than Q_{P1}, thus when expressed as a percentage, Q_{P2} minus Q_{P1} will be negative, and hence the percentage will be negative. When we put these values into our equation, then we get:

$$\text{CPED} = \frac{\%\Delta D_P}{\%\Delta P_C} = \frac{^{-}ve}{^{+}ve} = {}^{-}ve$$

Where:

$\%\Delta D_P$ = percentage change in quantity demand of bus services

$\%\Delta P_C$ = percentage change in the price of rail services

^{-}ve = a negative value

To summarise, substitute products will produce a positive value for the cross price elasticity, i.e. as the price of one increases, the demand for the other increases. Complementary goods however will produce a negative cross price elasticity of demand i.e. as the price of one increases, the demand for the other decreases.

3. INCOME ELASTICITY

Income elasticity of demand is a measure of the responsiveness of demand to changes in income. As incomes in real terms have been increasing (and continue to increase) over time, income elasticity tells us those markets that we may expect to see an increase in the future and those markets that perhaps, if other things remain equal, we may expect to see decline in the future.

Income elasticity of demand is calculated as:

$$\text{Income elasticity} = \frac{\text{percentage change in quantity demanded}}{\text{percentage change in income}}$$

Or using a shortened formula:

$$\text{YED} = \frac{\%\Delta D}{\%\Delta Y}$$

Where:

YED	=	income elasticity of demand
%	=	percentage
Δ	=	change (represented by the Greek letter 'delta')
D	=	quantity demanded of the good or service
Y	=	income

Note that income is always represented in Economics by the letter Y (and price is always represented by the letter P). Hence, YED stands for the income elasticity of demand. It should be noted that when we refer to income elasticity, what we are actually referring to is not the total income of consumers but rather what is known as their disposable income. This is income net of personal taxes, such as PAYE[8] and national insurance, and hence is the amount of income available for spending and saving.

As highlighted in Unit 1, transport is defined as a normal good, in that more is demanded at higher levels of income. It should follow therefore that the income elasticity of demand should be positive for transport. As concerning individual modes however, whilst the car, the train and air travel, all of which are used by higher income groups, would be considered to be normal goods, this does not apply to all modes of transport or to all situations. The bus for example may be considered as an inferior good, in that demand will fall for this form of public transport as incomes rise.

[8] PAYE- pay as you earn, the main form of income tax in the UK.

Does this mean therefore that because real incomes have been rising over time (and continue to rise in real terms), the demand for bus services will disappear altogether? Whilst there are many products in the past that have now disappeared from the market due to this factor, this is most unlikely for bus travel, and we maybe can distinguish between short and longer term income elasticities of demand. In the short term, income elasticity is negative, and hence as individuals increase their real incomes they will use other forms of transport, most notably the private car. For this to continue into the longer term however, all other things would have to remain equal. With particular reference to the bus, it is extremely unlikely that this will be the case, as issues such as pollution, land use, safety and so on become more acute. Hence in the longer term the income elasticity of demand for bus services is likely to become less elastic (less negative), and may even become positive.

Finally, it is worth highlighting that not only can income elasticity change over time but also that there are limits on short term income elasticity because travel is limited by the amount of time available. It has been suggested that although business and leisure travel increases with income, there comes a point when the demand curve flattens out or even begins to fall because limits are placed on the time available for travel.

4. SERVICE ELASTICITY

Under the title of elasticity of demand, we could technically calculate an elasticity of demand for all of our determinants of demand, however in the field of transport it is also worth highlighting service elasticity of demand. This is the measure of the effect of changes in service standards on the demand for an operator's transport facility, and can be calculated as:

$$\text{Service elasticity} = \frac{\text{percentage change in quantity demanded}}{\text{percentage change in quality of service}}$$

Or in short hand form as:

$$\text{SED} = \frac{\%\Delta D}{\%\Delta QS}$$

Where:

SED	=	service elasticity of demand
%	=	percentage
Δ	=	change (represented by the Greek letter 'delta')
D	=	quantity demanded of the good or service
QS	=	quality of service

This form of elasticity effectively embraces all influences on demand not covered by the previous types of elasticity of demand. It includes many factors affecting demand from a

change in frequency of service to a change in restaurant facilities. The example of a bus journey would include possible service elasticity factors such as the reliability / timekeeping of the service, the availability of information, the state of waiting facilities, the quality characteristics of the journey (e.g. noise, smoking, congestion), the attitude of the driver and the quality of the vehicle.

Note however that this elasticity of demand is quite different in form from the others highlighted in this section. Price elasticity for example tells the producer the extent to which consumers are price sensitive. Cross price elasticity gives a measure of the extent to which consumers are sensitive to the change in prices of other goods and services, i.e. substitutes and complements. Income elasticity highlights the extent to which consumers will change their behaviour with changes (normally rises) in income. What exactly however does 'service' elasticity tell us? If we had a figure of say 0.6, what exactly does this mean? Well it means that an improvement in quality will cause an increase in the level of demand, although every one percent increase in quality will only produce a 0.6% increase in quantity demanded. Note still that this means very little, as we would need information upon the costs of the increase in quality and the gains in revenue they would produce. Note finally that whilst the first three elasticity measures are absolutes, this last measure is probably better viewed as a relative. For example, it would help to address the issue of what factors (quality improvements) would have the largest impact of demand, as an example frequency versus reliability versus punctuality and so on, nevertheless we would still need cost and revenue information of each potential improvement. In other words, such 'elasticities' are more likely to be in the form of forecasts of the potential improvements to quality that such changes would bring about.

We finish of this unit by examining the effect of a change in the conditions of supply on the market price and quantity if there is a substantial variation in the price elasticity of demand for the good or service. If for example we know the price elasticity of demand for a product, we can predict the effect on price and quantity of a shift in the supply curve for that product. Figure 2 shows the effect of a shift in supply with two quite different demand curves (D_I and D_E).

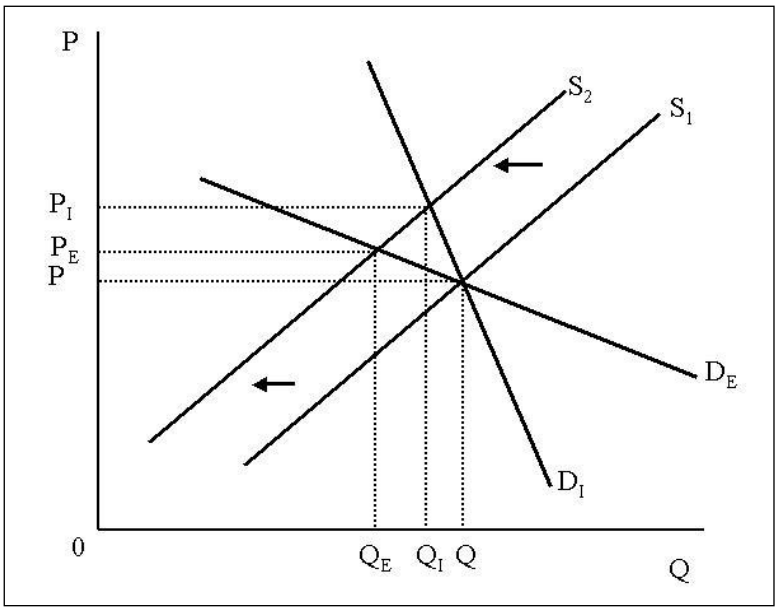


Figure 2.10: The effect in the market of a change in one of the conditions of supply on elastic and inelastic demand

This diagram can at first look quite messy. If we begin however with the market in equilibrium at price P and quantity Q, as represented by the supply curve S1, and then examine what happens if there is a change in one of the conditions of supply. Say for example there was an increase in the costs of production due to a pay settlement well above the rate of inflation. This would shift the supply curve from S_1 to S_2 (a decrease in supply). If we firstly examine the impact this would have if the demand for the good was relatively price inelastic. This is represented by the demand curve labelled D_I. The new market price would be given by price P_I and the quantity traded by Q_I. Note in this case the increase in costs has had little impact on the quantity traded and that most of the increase has been 'passed on' to the consumer in the form of higher prices.

If on the other hand demand for the good was relatively price elastic, then this would be represented by the demand curve D_E. In this case, the biggest impact of the increase in the costs of production would be a large decline in the quantity traded, as shown by Q_E, and subsequently a large decline in the size of the industry. In other words, most of the impact of the cost increase has been 'absorbed' by producers through reduced production, with only a relatively small increase in the price (to P_E).

This is one of the main reasons, if not the major nub of the argument, in favour of competitive industries. Under such a scenario, the firm faces a relatively elastic demand curve (due to the closeness and high number of available substitutes). As we have seen, this in turn produces strong downward pressure on costs, and hence such industries may be expected to be more 'efficient' than industries in which there is less competition. It may also partly explain one of the major problems with Railtrack, the former UK railway's

infrastructure company. Railtrack had a monopoly on the provision of infrastructure to passenger and freight operating companies, however its biggest problem was a considerable spiralling of its costs. These had been estimated by leading railway journalist Roger Ford to have increased in real terms by a factor of 2.5 under the management of the privatised infrastructure company[9]. Incompetence and the whole structuring of the company may have been other reasons, but these other issues are way outside the scope of this module!

Self-Assessment Exercises

Exercise 2.1

The following[10] is a working example of elasticity using pricing on the buses. Read it through and try to answer the questions at the bottom.

ANY MORE FARES? - Pricing on the buses

Imagine that a local bus company is faced with increased costs and fears that it will make a loss. What should it do? The most likely response of the company will be to raise its fares. But this may be the wrong policy, especially if existing services are under utilised. To help it decide what to do, it commissions a survey to estimate passenger demand at three different fares: the current fare of 10p per mile, a higher fare of 12p and a lower fare of 8p. The results of the survey are shown in the first two columns of the table.

Demand turns out to be elastic. This is because of the existence of alternative means of transport. As a result of the elastic demand, total revenue can be increased by reducing the fare from the current 10p to 8p. Revenue rises from £400,000 to £480,000 per annum. But what will happen to the company's profits? Its profit is the difference between the total revenue from passengers and its total costs of operating the service. If buses are currently under utilised, it is likely that the extra passengers can be carried without the need for extra buses, and hence no extra cost. At a fare of 10p, the old profit was £40,000 (£400,000 - £360,000). After the increase in costs, a 10p fare now gives a loss of £40,000 (£400,000 - £440,000). By raising the fare to 12p, the loss is increased to £80,000. But by lowering the fare to 8p, a profit of £40,000 can again be made.

[9] Although Cowie & Loynes (2007) describe Ford's 2.5 value as a 'headline' figure, more applicable to new projects. Cowie & Loynes (2007) nevertheless estimated cost escalations to be in the order of 40% for maintenance and renewal.

[10] Adapted from Sloman J (2000), *Economics,* p57.

Fare (pence per mile)	Estimated demand (passenger miles per year: millions)	Total revenue (£ per year)	Old total cost (£ per year)	New total cost (£ per year)
8	6	480,000	360,000	440,000
10	4	400,000	360,000	440,000
12	3	360,000	360,000	440,000

1. Estimate the price elasticity of demand for a price fall from 10p to 8p and for a price rise from 10p and 12p.
2. Was the 10p fare the best fare originally?
3. The company considers lowering the fare to 6p, and estimates that demand will be 8.5 million passenger miles. To meet this demand however, it will have to put on extra buses. Advise the company how it should make this decision whether or not to reduce the price. (HINT: give the company a maximum figure for the increase in costs resulting from the extras buses).

Exercise 2.2

Consider at a general level the government's taxation policy. You may be under the false illusion that commodities such as alcohol and tobacco are heavily taxed for 'health' reasons. You should now re-consider this view, and compare through two illustrations of what would happen to the tax take (the total revenue generated from a tax) if the government was to increase the tax on an elastic and an inelastic good.

Exercise 2.2[11]

This is a totally artificial exercise, however it is designed to try to get you to think about own price, cross price and income elasticities. Presented below are some completely hypothetical passenger figures for public transport services in a hypothetical city somewhere near you!

Transport Mode:	Rail	Bus	Underground	Total
Annual usage (millions):	38	90	23	151

For this hypothetical public transport market, the following elasticities apply:

		Rail	Bus	Underground
Income elasticity of demand:		0.41	-0.50	0.32
			Price	
Own & cross price elasticities		Rail	Bus	Underground
	Rail	-0.45	-0.40	-0.30
Quantity	Bus	0.08	-0.40	0.10
	Underground	0.02	0.05	-0.20

Note: modes listed on rows relates to the quantity change in demand, those listed in columns relates to change in price

Using all of these values you should be able to answer the following questions – as a side note, if you have the necessary skills you may find a spreadsheet useful to assist with this exercise.

a. If there is a 5 per cent rise in income, what would be the new daily modal splits and the new total daily usage?
b. Using your answer for the new total daily usage from part (a), what is the overall income elasticity to travel?
c. How does your answer from part (b) compare with the results presented in Case Study 3.1 in Chapter 3 and what might be the reason for any such differences? (Hint: you will need to calculate a rough elasticity from the values presented in the case study)
d. Calculate the effect on modal splits and the new monthly usage of the impact of the following factors (each should be considered on its own) and from your

[11] Copied directly from Cowie (2010), but don't anyone tell him!

answers highlight which modal fare has the largest impact on the overall demand for travel in this city.

i. a 15 per cent increase in the level of rail fares
ii. a 15 per cent increase in the level of bus fares
iii. a 15 per cent increase in the level of underground fares

e. What might be expected to happen to the cross price elasticity of the train across all other modes if the level of rail travel was to significantly increase? Why would this happen?
f. Roughly speaking, why have we got the answers that we have got for part c and what does this underline with regard to own and cross price elasticities of public transport services?

UNIT SUMMARY

In this unit we have examined the elasticity of demand of consumers. The four types of elasticity of demand relevant to the transport sector have been outlined, namely price elasticity, cross elasticity, income elasticity and service elasticity. You will now be able to calculate and graph the elasticity of demand. We have also seen how the price elasticity of demand can be linked to consumer expenditure.

UNIT THREE
FACTORS BEHIND SUPPLY – TRANSPORT COSTS

INTRODUCTION

In Unit 2 we examined the factors behind demand, in particular the elasticity of demand. In this unit we examine the factors behind supply, namely transport costs and the behaviour of costs as the level of output varies, as this has a major impact on how transport services are provided in the market. In order to do this we will have to formally define time in Economics, examine the behaviour of output as more factors of production are added, and then finally consider the cost implications from that analysis. In the following unit, we will then put all of our ideas together concerning supply and demand to examine the theory of the firm.

RECOMMENDED READING

The following chapters, from four references, are recommended reading for this Unit:

Cowie, J. (2010). *The Economics of Transport*, Chapter 5, Transport Costs

Cole S (1998), *Applied Transport Economics*, Chapter 4. Costing.

Button K J (1993), *Transport Economics*, Chapter 4. The direct costs of transport.

White P (2008), *Public Transport: its planning, management and operation*, Chapter 6. Costing and cost allocation methods.

LEARNING OUTCOMES

Once you have completed this unit you should be able to:

- Understand the concept of time as defined by the discipline of Economics;
- Understand the nature and type of transport costs;
- Understand the difference between the short and the long run in economics
- Explain the concepts of technical, cost and allocative efficiency
- Compare and contrast the differences between the law of diminishing marginal returns, returns to scale and economies of scale

THE ECONOMIST'S DEFINITION OF TIME

We begin our examination of costs by firstly considering the issue of time. We need to examine time and devise formal definitions for what is a relatively 'short' period time and what is a relatively 'long' period of time. The reason for doing this is that costs may behave quite differently depending upon whether we are considering a relatively short period of time or a relatively long period of time.

In unit one the idea of the factors of production was introduced. To briefly re-cap, the inputs into any production process can simply be divided into three categories, land and raw materials, labour and capital. The extent to which these factors of production vary in order to produce a different level of output gives us a formal definition of time, in particular:

The Short Run

The short run was first explained to me as about 50 metres, however in economics the short run is formally defined as that period of time during which at least one of the factors of production is fixed. This could in theory be any of the inputs, however normally relates to capital. The implication therefore is that variations in the level of output can only be achieved through variation in one or more of the other inputs, normally labour. This may be achieved through overtime working, employment of agency labour and so on.

The Long Run

In the long run, variations in the level of output can be achieved through variation of all of the inputs, thus we are not restricted to using only one of them. Capital can therefore be expanded to achieve such gains.

As regards an actual time period for the short and long runs, i.e. months, years and so on, this can vary considerably from industry to industry. In relatively labour intensive industries, i.e. those that employ a high degree of labour, such as the bus industry, the length of time in the short run can be probably measured in weeks. Using the bus industry as an example, a few weeks is roughly the length of time it would take to increase the capital stock of the firm i.e. purchase a second hand bus. Procurements of new buses however, will probably take considerably longer and would be measured in months. In more capital intensive industries, i.e. those that employ a relatively high degree of capital in the production process, such as the rail industry, the short run can probably last up to a number of years. Taking the rail industry, the gap between ordering new rolling stock and it actually being introduced into service will take a minimum of roughly two years and can take anything up to five. There is therefore no 'fixed' period of time as such with regard to the short and long runs, it will vary from industry to industry dependent upon how long it takes to vary all of the factors of production.

The Very Long Run

The final definition of time in economics is the very long run. In simple terms the very long run is that period of time where all factors of production are variable, including the level of technology.

THE PRODUCTION PROCESS

Production is the process whereby economic inputs or resources are brought together by entrepreneurs (or firms) to produce goods and services. This basic process is outlined in figure 3.1 below:

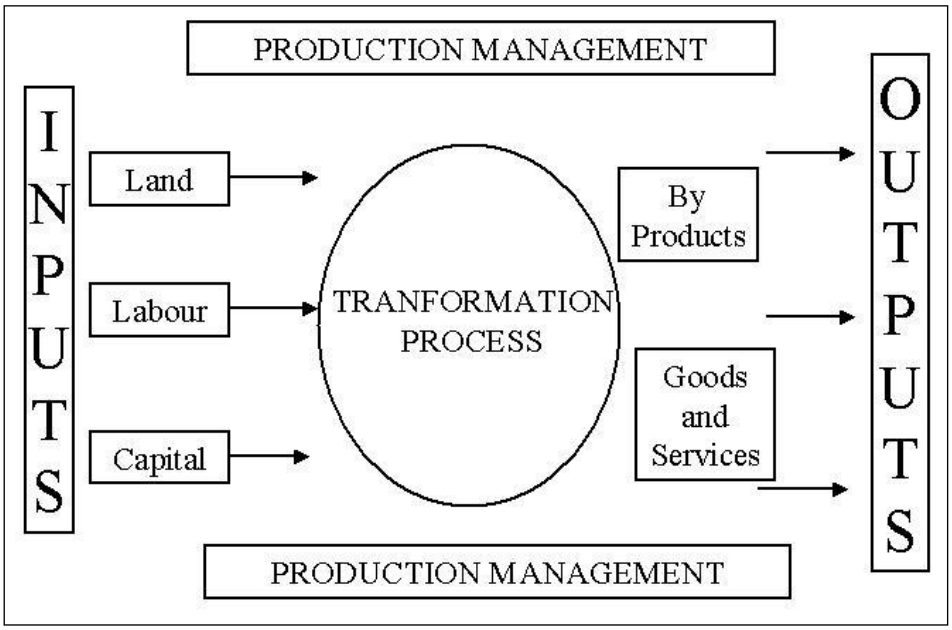


Figure 3.1: The Production Process

As highlighted by Figure 3.1, the production process consists of converting the inputs of Land (and Raw Materials), Labour and Capital, by way of a transformation process, into the outputs of Goods and Services and By Products. Goods and services are then sold for an economic return, with the aim being that the transformation process should have converted the inputs into something of a higher economic value. In other words, the revenue generated from the sale of the output should be greater than the payment for the inputs. Also shown in Figure 3.1 is 'By products'. Not all of the output of the transformation process is sold for an economic return, as the transformation process also results in the output of other factors such as wastage or pollution. These have to be counted in the outputs of the production process.

The transformation process is known as production, and the relationship between the level of inputs and the level of outputs known as the production function. This is given by the relationship shown below:

$$Q = f(A, L, K)$$

Where:

Q = quantity of output produced

F = 'some function of'

A = quantity of land and raw materials used in the production process

L = quantity of labour used in the production process

K = quantity of capital used in the production process

This equation simply states that the level of output is some function of land and raw materials, labour and capital. This inputs/outputs ratio is one of the bases for assessing whether a given operation can be described as being 'efficient' or not. 'Efficiency' as such has many different meanings to different people, and can encompass issues such as reliability, punctuality, pricing, costs and so on. Within economics however, we have three formal definitions of efficiency, and these are:

Technical efficiency – this relates to the outputs/inputs ratio, and refers to using the minimum level of inputs to produce the maximum level of outputs. This would be known as technically efficient if we could not achieve any more output from the level of inputs used.

Cost efficiency – cost efficiency arises as in order to produce a level of output there may be several ways of doing it – either use a high level of capital and low level of labour, or a high level of labour and a low level of capital. Both could be technically efficient (the inputs to outputs ratio). The key issue therefore is which is the lowest cost combination of the choices available? The one that is would be the cost efficient position.

Allocative efficiency – even within the study of economics, there is much confusion over the term 'allocative' efficiency, which is quite surprising given it is almost the holy grail of the discipline! Allocative efficiency however relates to usage. As many of the former Eastern European countries showed in the past, there is little point in producing goods and services that are technically efficient in the lowest cost combinations if no one wants them. This is a total wastage of resources and hence could never be considered as efficient.

PRODUCTION IN THE SHORT RUN

The most basic form of the production function is where only one input is varied to produce variations in the level of output i.e. short run production. The example in table 3.1 shows the basic relationship between adding a variable input (labour) to a fixed input (capital) to increase the level of output. The hypothetical example concerns the production of bus services, from where we employ zero labour units up to a maximum of nine, and output refers to thousands of vehicle kilometres[12].

Table 3.1: Variable Labour and Production of Bus Services

Unit	Output Produced (in thousands)	Average Product (thousands)	Marginal Product (thousands)
0	0.0	0.0	
1	0.0	0.0	0.0
2	3.0	1.5	3.0
3	7.5	2.5	4.5
4	14.0	3.5	6.5
5	22.5	4.5	8.5
6	30.0	5.0	7.5
7	35.0	5.0	5.0
8	36.0	4.5	1.0
9	31.5	3.5	-4.5

.

And this is graphed over the page in Figure 3.2:

[12] The issue of what actually constitutes the output of transport companies is an interesting one in itself. Transport is of course a service, and as such there is no actual physical output, but rather it is the passenger journey or the freight moved that is produced. In terms of measurable outputs however, most academic studies on the topic have used some measure of supply, such as vehicle kilometres, rather than some measure of demand, such as passenger journeys. Space is far too short as to go into the reasons why this is the case, but simply to say that a supply based measure better represents the service aspect of transport companies.

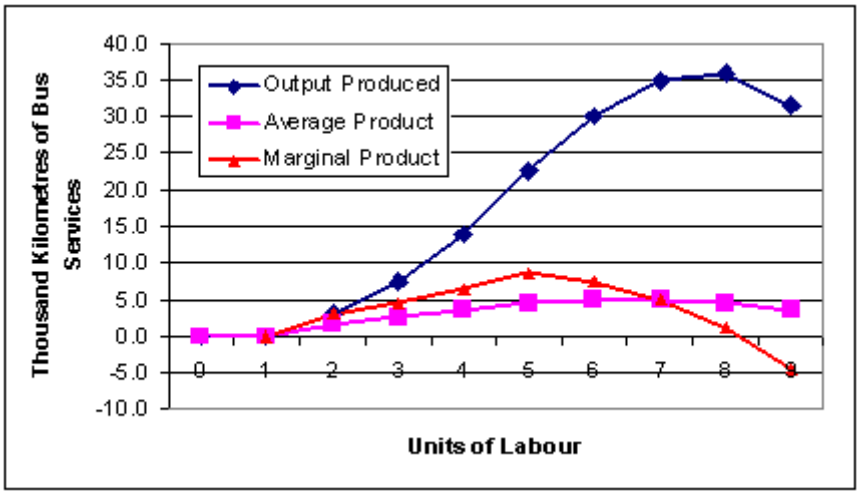


Figure 3.2: Total, Average and Marginal Products

This example is used to show the basic theoretical relationship between adding a single variable input to a fixed unit and the effect that this has upon the level of output. As all output variations are produced by the addition of the variable input, all changes in the level of output are ascribed to that input. Starting at zero, no output is produced, as the only factors of production that are employed are the fixed factors, and we need all factors to produce any output. Even after a single unit of the variable factor, labour, is employed still no output is produced. This is because by the time the single employee has opened up the bus station/depot, cleaned the facilities, carried out all of the admin and maintenance tasks and so on, there is no free time left to actually drive a bus!

When a second labour unit is employed however, then a degree of specialisation can occur – one can look after the bus station whilst the other drives the bus. As more units are employed, then more specialisation can occur and hence the level of output increases at an increasing rate. This is shown by an increase in both the average product and the marginal product figures in Table 3.1. This increasing increasing rate continues up to a point of five labour units, at which point the marginal product is at its maximum point. With the employment of a sixth person, although output still rises it is rising at a decreasing rate, hence although the average product is still rising, marginal product begins to fall. When we get to nine units however, the total output actually falls. This can be for a number of reasons, such as employees are now getting in each other's way (remember, all other inputs are fixed), too much chat etc etc.

In economics, this tailing off of the marginal product is known as the law of diminishing marginal returns. Peppers and Bails (1987) explain that the law of diminishing marginal returns states that the marginal product of the variable factor of production will eventually decline if enough of it is combined with the fixed factor. In other words, we cannot endlessly continue to add a variable input to a fixed factor and continually achieve increases in the level of output. At some point we will reach maximum capacity, where in order to increase output further we will have to increase the level of the fixed factor. As

we approach that maximum capacity point, then the marginal returns from our variable factor will decline.

We can also use this example to break down our production processes into three stages, known conveniently as Stage 1, Stage 2 and Stage 3 production. These are shown in Figure 3.3.

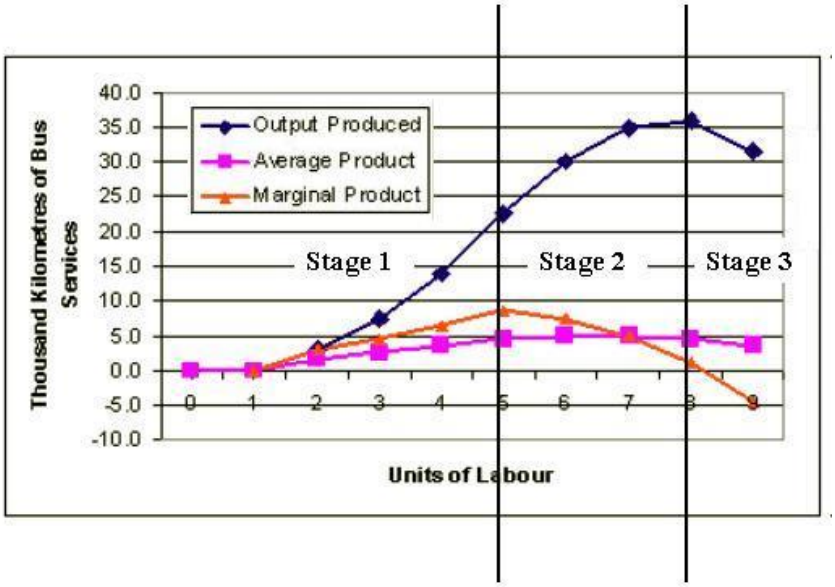


Figure 3.3: Stage 1, Stage 2 and Stage 3 production

You should be able to see from figure 3.3 that this division is based upon the marginal product, as this tells us the extent to which the rate of production growth is increasing or decreasing. In stage one, the marginal product is always increasing, in stage two it is decreasing but is still positive, and in stage 3 production marginal product becomes negative.

One final issue to consider before we leave short run production is that the marginal product curve cuts the average product curve at the latter's highest point i.e. at seven labour units. This is a simple mathematical concept. While we are adding extra units of labour and the total that they contribute in terms of output is above the average, then this will increase the average. If however the last person employed should add less to the total output than the average, this will pull the average down[13].

[13] Just to make this clear it may be worthwhile to consider a non-economic example. If there are three people in a room, ages 20, 30 and 40, then the average age in the room would be 30. If a fourth person walks in, say they are 22, the marginal age would be 22 and the average age of the four would be 28 i.e. because the last person's age was less than the average they brought down the average age. Alternatively, if this person had been 38, then because this is above the average age it would pull the average up to 32.

The Theory of Costs

CLASSIFYING COSTS

There are a number of possible ways of classifying costs in a transport operation, which enable analysis of the characteristics of costs as output varies and of the actual costs themselves. Some of these you may study in different modules during the course. Within transport economics however, there is only one way of classifying costs, and that is into fixed and variable. As the names perhaps suggest, a fixed cost is one that does not vary with the level of output, whilst a variable cost does. Although we have simply slipped this in as almost a by-line, this balance between fixed and variable costs can and does have considerable implications on the supply of transport services, and these we will examine in the next unit.

Fixed costs are input costs that are often classified as 'indivisibles', such as rents and rates that relate to offices, depots, stations and so on, management salaries and costs, administrative expenses such as telephones etc, that must be paid for even if there is a zero level of production. Variable costs on the other hand relate to the direct expense of providing the output, such as wages and salaries, power, heat and light and so on. Many costs however are semi-variable, that is they partially vary in relation to output. Wages and salaries for example technically would actually be a semi-variable cost. In terms of this analysis however, we will consider all semi-variable costs to be wholly variable. Examples of transport costs are given over the page in Table 3.2.

Table 3.2: Transport Operators Main Fixed and Variable Costs

Cost	Content	Type
Traffic operating costs	Drivers, conductors and supervisory staff, ticket equipment, uniforms, miscellaneous supplies	Variable
Fuel and power	Self explanatory	Variable
Servicing, repairs and maintenance	Staff costs, materials and spares, maintenance of buildings, machinery etc.	Variable
Management costs and welfare	Administration, insurance, welfare and medical facilities, etc.	Fixed
Rent, leases and rates on buildings	Station terminals, office buildings, bus stations, depots and garages, service works	Fixed
Depreciation	Wear and tear on equipment, generally the cost of using the capital equipment	Technically semi-variable, but in the short run normally treated as fixed

All of these hopefully are straightforward except the last one, depreciation, although this is a cost we have already encountered on the Public Transport module. Depreciation has two definitions, an economic definition and an accounting definition. In economic terms, depreciation is the reduction in the value of an asset to the firm through actually using it. If we use a bus for example, there will come a point in time where it will be cheaper in the longer term to replace that bus rather than continue to 'patch it up' and keep it running. How depreciation is handled in the accounts however, is either by writing off a set amount every year, or writing off a percentage of the value every year. Say for example a new bus costs a firm £100,000 and is expected to last nine years and has a scrap value of £10,000. The company would therefore write off £10,000 a year, hence in nine years it would be shown in the books as having a value of £10,000 i.e. it's scrap value. By the second method, writing off about 22½% of the book value each year will achieve the same purpose, but note that under the 'reducing balance' method higher values are written off in the earlier years. Irrespective of what method is used however, neither of these methods takes into account usage. For example, under the first measure £10,000 will be written off the value of a bus each year irrespective of the extent to which that bus

is used. In this sense therefore, it is a fixed cost, as it is one that does not vary with the level of output produced. In the longer term however, if that bus was little used it would retain its economic value, and hence would still be useable to the firm long after it had been written off in the books. There are numerous examples of transport assets that have long been written off but are still in use today, such as the Severn Rail Tunnel and the Forth Bridge.

In order to add further confusion however, say a single bus operator wished to double output, what would it do? One of the things it would have to do is to buy another bus, hence the annual depreciation charge would increase. In that sense therefore, it is a semi-variable cost. The issue of deprecation can be very confusing (as you have just seen!), but as long as you are aware exactly what depreciation is and the difference between the economic value and how depreciation is treated in the company's books that is all that is really important. We can side step much of this debate anyway when we consider the behaviour of costs in the short run, as during the period of the short run at least one of our factors of production (usually capital) is fixed. In the following example therefore, the number of buses is fixed, and consequently the depreciation charge is fixed.

Purely for illustrative purposes, if we expand the previous example of the production of bus services in the short run and include cost data. If we simply assume that our fixed costs are £80,000 and the variable factor, Labour, costs £30,000 per unit, then we can expand Table 3.1 to include this cost information:

Table 3.3: Variable and Fixed Costs of Short-run Production of Bus Services

Labour Units	Output Produced (000s)	Average Product (000s)	Marginal Product (000s)	Total Fixed Costs	Total Variable Costs	Total Costs	Average Total Costs	Marginal Costs[1]
0	-	-	-	80000	-	80000	-	-
1	1	1.0	1	80000	30000	110000	110.00	30.00
2	7	3.5	6	80000	60000	140000	20.00	5.00
3	18	6.0	11	80000	90000	170000	9.44	2.73
4	26	6.5	8	80000	120000	200000	7.69	3.75
5	32	6.4	6	80000	150000	230000	7.19	5.00
6	37	6.2	5	80000	180000	260000	7.03	6.00
7	40	5.7	3	80000	210000	290000	7.25	10.00
8	42	5.3	2	80000	240000	320000	7.62	15.00
9	41	4.6	-1	80000	270000	350000	8.54	-

Note: 1. Technically marginal cost should be placed in between the rows, hence all of these should be moved down half a line.

The first four columns are a simple lift from Table 3.1, and to this we have added five new columns to include our cost data. Total Fixed Costs, as outlined above, do not vary with the level of output hence remain fixed at £80k irrespective of the level of output that is produced. Total Variable Costs are simply the number of labour units employed times the cost of each unit, and total costs is the addition of fixed and variable costs. Average total cost is just the total cost divided by the output, and the marginal cost is the cost of the last unit produced. If for example we move from one to two labour units, then the output increases by 6,000 units, and costs rise by £30,000. Therefore, the marginal cost is £5. If you have not got all of that, then you should work through some of the lines yourself to make sure that you have the basic ideas.

If we graph Table 3.3, we can then examine the shape of the short run average and marginal cost curves. This is done in Figure 3.4.

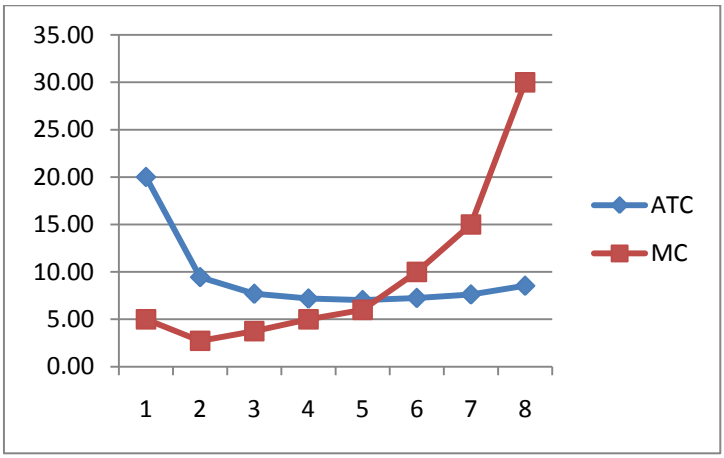


Figure 3.4: Short-run Average and Marginal cost Curves

Although not clearly visible from this example, the average cost curve falls and then rises i.e. there is an optimal level of production in the short run in terms of the average cost. The marginal cost curve cuts the average cost curve at the latter's lowest point, for the same reasons that the marginal product curve cut the average product curve at its highest point in the previous example. To give a more textbook 'stylised' example, further graphs are given below:

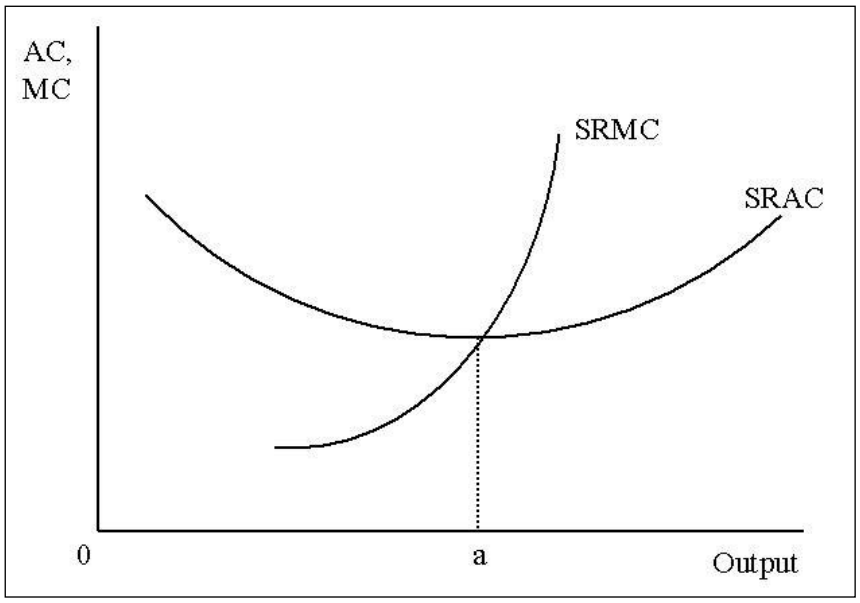


Figure 3.5: Short run Average and Marginal Cost Curves

This is the stylised version of the short run marginal and average cost curves. As more of the variable input is added, average costs at first fall, and are then minimised at the optimal output level of production at point a. Beyond that point, average costs then rise.

Having outlined the basic shape of these curves, what is important is why is the short run average cost curve U-shaped. The simple answer is due to the law of diminishing marginal returns. In other words, as defined above, the marginal product of the variable factor of production will eventually decline if enough of it is combined with the fixed factor. Consequently, the rate of decline in average costs will reduce, and eventually average costs will rise. That's fine for textbook economics, but are there any 'common sense' reasons as to why this should be the case?

If one of our inputs in the production process is fixed, let's say the number of buses, then this should suggest that there is an optimal production level, i.e. an ideal level. Any production level that is under that point will lead to the under-utilisation of the fixed resources. In turn the fixed costs associated with employing these factors will be spread over a relatively low level of output hence the average cost will be relatively high. Any level that is over the optimum however, whilst spreading fixed costs over a high level of output, will lead to the over-utilisation of resources. As we can only alter the level of production by varying the level of the variable factor, in this case labour, then in order to increase labour above the optimal level this may involve paying over-time rates, hiring agency labour at a higher cost and so on. Therefore, there are both theoretical reasons and cost reasons as to why the short run average cost curve is U-shaped.

COSTS IN THE LONG RUN

The first point about costs in the long run is that there is no division between 'fixed' costs and 'variable' costs, because in the long run as all inputs are variable then all costs are variable. To briefly re-cap, whilst our firm may be planning a new production facility or entering a new market, it is operating in the long run. Once however it builds the new factory or sets up the new depot to service the new market, it is then operating in the short run (because at least one of the factors of production is fixed).

How would we expect costs to behave in the long run therefore? We begin to examine this by again briefly looking at the behaviour of production as more inputs are added. Figure 3.6 therefore illustrates a long run production function.

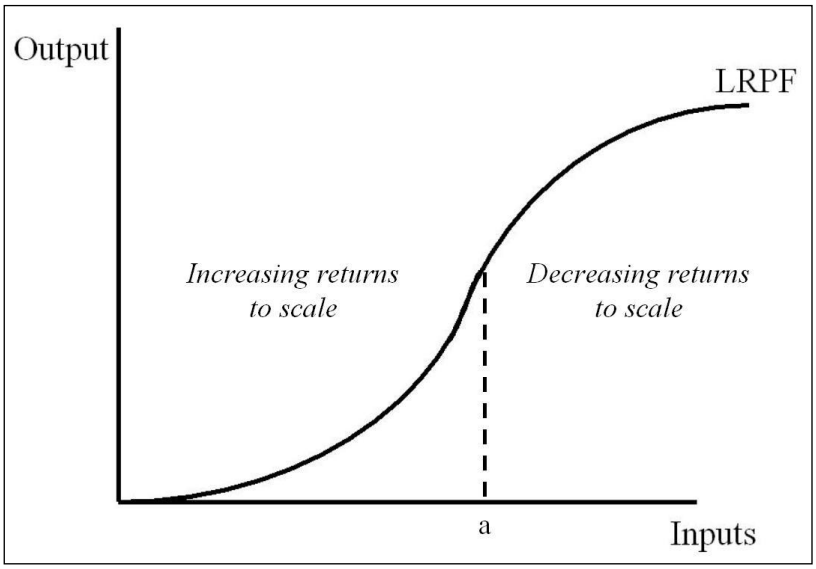


Figure 3.6: The long run production function

The long run production function is S shaped in nature. At first there are large gains when firm size increases – the relative percentage gain in output is greater than the relative percentage increase in inputs. Note as well that this effect as such is increasing as firm size is increasing. In other words, increases in productivity are getting larger as firm size gets larger. These increasing gains continue up to point a on Figure 3.6.

The effect of firm size on the level of productivity is known as returns to scale, and where these effects are becoming larger as output grows, this is known as increasing returns to scale. Therefore the area to the left of point a is the area of increasing returns to scale. Once we are past point a however, although at first the proportionate gains in output are still larger than the proportionate increases in inputs, these productivity gains are not as large as before, in other words there are decreasing returns to scale. This decline continues until increasing firm size has very little effect on the level of output. This is shown in Figure 3.4 in the area to the right of point a.

Why does this occur? Note that as we are dealing with the long run, the correct answer is NOT because of the law of diminishing returns, as this only applies where at least one of our inputs is fixed, i.e. in the short run. The main reasons for this general shape are outlined below.

SOURCES OF INCREASING RETURNS TO SCALE

Specialisation of Labour – larger firms allow more specialisation of the workforce to occur. If we consider this at the most basic level, in a one-person sole trader business then the owner has to undertake all of the tasks involved in running the company, they therefore become a 'jack of all trades and master of none'. As the firm increases in size however, then more labour can be employed in specialised tasks and thus become more proficient at those tasks. Consequently we would expect productivity to increase. Note also that as this is the long run, there is potentially no ceiling on this effect.

Scheduling of inputs – similar to the specialisation of labour, except scheduling of inputs refers to all of the inputs, and this factor is particularly prevalent in the transport industries. As firm size increases, there exists greater flexibility in how the inputs can be combined and hence better utilisation of all of the inputs may be expected i.e. higher productivity.

Capital inputs – this concerns a number of issues that we have broadly grouped together under 'capital inputs'. Firstly, some capital inputs can be very expensive, such as research and development and investment. Only larger firms can afford to spend on these issues, but they only do so on the basis that this will lead to improved efficiency (i.e. higher productivity) in the longer term. The second issue under this title is again to do with specialisation. Relating back to our one-person sole trader, then it may well be that the company vehicles consist of a single solitary van. This van will have to fulfil all of the transport requirements of the company, some of which it may be better suited to than others. As firm size increases however, then the company fleet can be increased not only in size but also in the degree of speciality, hence more suitable vehicles can be used for more suitable tasks. This should lead to higher productivity. Thirdly, it may make more sense for larger companies to carry a larger number of spares and maintenance facilities, hence downtime of capital equipment as such should be reduced and consequently higher productivity achieved[14].

Indivisibilities. The standard example of an indivisibility is a telephone line. When setting up in business, a company will need to install and rent a phone line. With small

[14] It should be noted that even with 'just-in-time' production methods and the contracting out of maintenance and servicing facilities (as has happened with many bus companies), we may still expect certain advantages with regard to this issue. Larger companies for example may be anticipated to have more 'clout' with suppliers of such services in the form of the latter responding more quickly to service requests. We may expect capital downtime as such therefore to be lower.

expansions in size, there will probably be no need to install and rent a second phone line, hence this 'input' as such is spread over a larger output. There will obviously come a point where a second phone line will have to be installed, but this would only be done if it was advantageous to do so e.g. if it improved 'productivity'. Another example of an indivisibility of course is our sole trader's van. As firm size increases, the van's overall utilisation would improve.

Hopefully these sources of increasing returns to scale all seem reasonably clear, but what may cause this increase in productivity to tail-off, i.e. sources of decreasing returns to scale? This is particularly prevalent in the long run as there is no upper constraint in the form of a fixed input.

SOURCES OF DECREASING RETURNS TO SCALE

Loss of Control – as firm size increases, there is a loss of control over the whole organisation. This loss of control decreases overall productivity. The company no longer becomes a lean mean machine. Under loss of control it is also worth highlighting the concept of 'X-inefficiency', which was originally devised by Harvey Leibenstein (1966). In simple terms, x-inefficiency relates to general management slack, and large (publicly owned) organisations are said to suffer from this concept. In less academic terms, in larger organisations there may be a loss of the sense of the individual, and more opportunities for general slack working practices, hence leading to lower productivity levels.

Geographical location. Particularly prevalent in the bus industry, but true of other transport industries, when a firm first sets up its business it will probably be on or near to the optimal location. Increasing size in the longer term means building other production facilities, depots whatever, and these will not be on the best location, hence productivity is reduced.

Administration procedures – large firms often require many more layers of middle and upper management, plus administration procedures in order to control costs and processes within the organisation. This is commonly known as 'bureaucracy', however in this situation we are not referring to what is commonly known as 'red tape', as we consider that in the next section. More specifically, we are referring to the time dimension that such 'form filling' requires and hence the opportunity cost of such form filling is the distraction of employees from the production process. When measured in terms of overall output therefore, it requires a relatively higher number of employees to produce a higher level of output.

COSTS IN THE LONG RUN

Having considered production in the long run, it should be fairly clear how we would expect costs to behave in the long run. These are graphed in Figure 3.7.

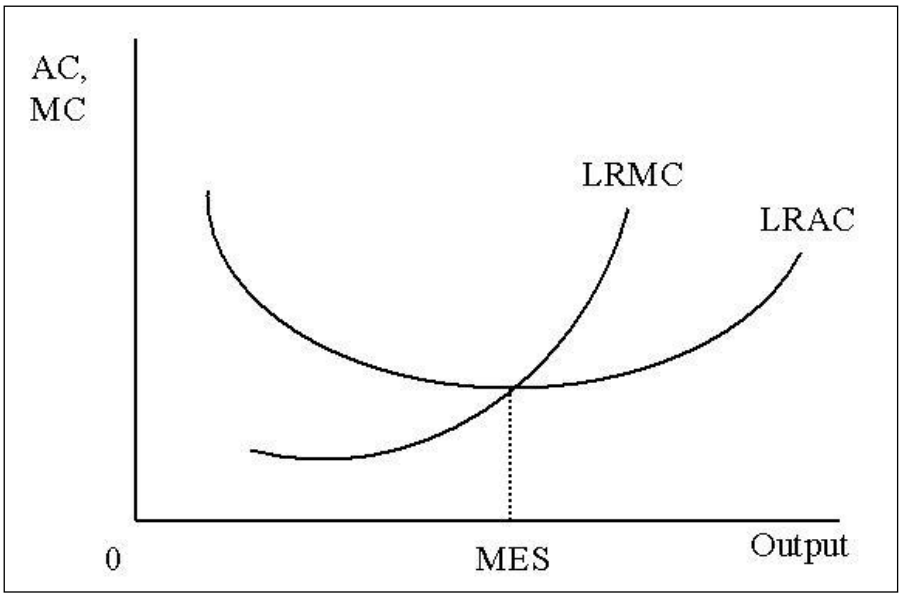


Figure 3.7: Long run average and marginal costs

The first point to notice from Figure 3.5 is that average and marginal costs behave exactly the same in the long run as they do in the short run. As can be seen from Figure 3.7, average costs at first fall as firm size (as measured by output) increases. There comes a point however, known as the minimum efficiency scale (MES), where this trend reverses and average costs then increase with firm size. The optimum level of production therefore is at the MES point, as this is where long run average costs are at their lowest. When the average cost is falling as firm size is increasing, this is known as economies of scale. Very often however, these are incorrectly termed increasing returns to scale. You should be clear on the difference in the two concepts, as returns to scale relates to production units while economies of scale relates to production costs. When average costs increase as firm size grows, this is known as diseconomies of scale, or again often incorrectly referred to as decreasing returns to scale.

Why should long run average costs first fall and then rise as output (and hence firm size) increases? Hopefully, you should already know the first source of these economies of scale.

SOURCES OF ECONOMIES OF SCALE

Increasing returns to scale – as we have seen above, at first as firm size increases it encounters increasing returns to scale i.e. ever higher levels of productivity. This increased productivity therefore means that relatively lower levels of the inputs need to be employed to produced higher levels of output, hence the average costs associated with these inputs per unit of output falls.

Bulk buying – larger firms can normally obtain some form of discount for buying capital equipment and supplies in larger numbers, and hence we would expect average costs to be lower for larger firms.

High cost inputs – as well as a source of returns to scale, high cost inputs are also a source of economies of scale. Some inputs can be very expensive, such as research and development and investment. Only larger firms can afford to spend on these issues, but they only do so on the basis that this will lead to improved efficiency in the longer term and hence reduced average cost. Advertising is also sometimes cited as a high cost input, but this is debatable if this really leads to economies of scale. One example of where it probably does however is in the Cola market. By advertising extensively, firms may seek to increase the size of the market and hence allow them to take advantage of other economies of scale through increased production to service that larger market.

Financial economies – larger firms are normally better placed to secure additional finance as they can offer greater security. Interest rates therefore may be lower as there is a lower risk involved, and hence average costs are lower for larger firms.

SOURCES OF DISECONOMIES OF SCALE

Decreasing returns to scale. As highlighted above, there are a number of sources of technical inefficiencies that lead to reduced productivity for larger firms, thus the average output per unit of input falls. In order to produce higher levels of output therefore, relatively higher numbers of inputs need to be employed, and this adds to costs causing diseconomies of scale.

Red Tape – as noted above, larger firms often require many more layers of middle and upper management, plus administration procedures in order to control costs and processes within the organisation. This is what is commonly known as ‘red tape’, and the actual cost of this added administration burden will add to costs, hence increasing average cost per unit and is thus another source of diseconomies of scale.

Transportation costs – if a firm increases long run output by expanding the size of current production facilities, then in order to serve new markets the company will have to transport the output further, hence increasing average cost.

SHORT AND LONG RUN AVERAGE COSTS

Before we examine the practical implications of what has been outlined, we briefly consider the relationship between the short and the long run average and marginal cost curves. This is shown in Figure 3.8. This is useful to further underline the relationship between the two concepts of the short and long run.

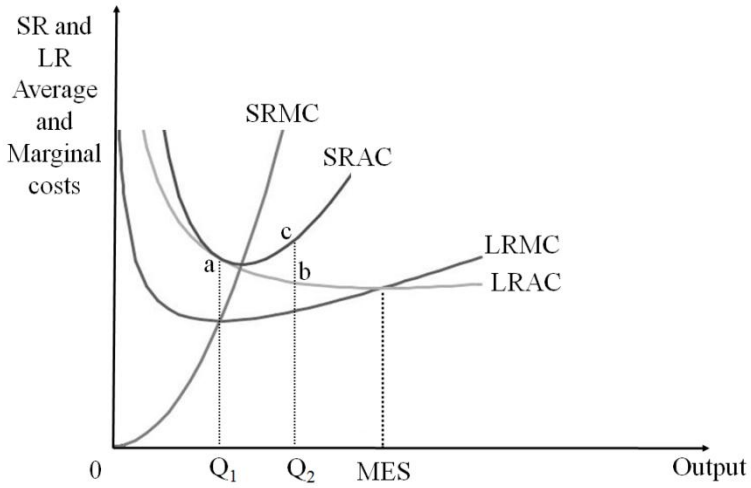


Figure 3.8: Long and short run average and marginal cost curves
Source: Cowie (2010)

As should be seen from figure 3.8, each short run average cost curve is tangential to the long run average cost curve, thus there will be differences if increases in output are produced by holding one of the factors of production fixed, i.e. the short run, or if the firm is able to increase the level of all of the factors of production, i.e. the long run. Beginning at point a, if the firm were to increase production from Q1 to Q2 in the short run its average cost would follow the short run average cost curve, and end up at point c. If however it could expand production from Q1 to Q2 in the long run, it would end up at point b. To re-inforce the point, b is lower than c because hat is because when operating in the long run the firm effectively has a blank sheet of paper, and hence can plan for the lowest cost level of production. Once one of the factors is fixed however, then it is more confined and it has to work around this constraint, which will incur a higher average cost. Note also the relationship between the marginal cost curves, and make sure that you understand the difference between the long and short run marginal costs. Note finally that it took me ages to draw this diagram!

Unit 3 Self-assessment exercise

This exercise concerns the provision of rail services, and the task is comparatively straight forward. In addition to cost information you are also given price information. You therefore have to fill all the blanks in on the table and thereby find the profit maximising position. You should round all figures to one decimal place, and will need the following information:

Fixed Costs: 100000 Price of a variable factor: 50000

Unit	Output (000)	AP	MP	TFC	TVC	TC	ATC	MC	AR	MR	TR	Profit	Price
0	0												
1	50												1.5
2	110												1.4
3	180												1.3
4	260												1.2
5	350												1.1
6	420												1.0
7	480												0.9
8	530												0.8
9	570												0.7
10	590												0.6

AP = average product MP = marginal product TFC = total fixed costs TVC = total variable costs TC = total costs
ATC = average total costs MC = marginal cost AR = average revenue MR = marginal revenue TR = total revenue

1. What is the most technically efficient level of output?
2. What is the most cost efficient level of output?
3. What is the level of output that the firm should produce at?

Case Study – Economies of Scale in the Rail Industry

The general view of economies of scale within the rail industry is that due to the fact that the industry employs a high level of capital equipment, economies of scale are hugely significant and hence company size needs to be large in order to capture all of these effects. Capital costs tend to be fixed and hence a large company size is required to spread these costs over a large output range and hence reduce the average costs per unit. This in the past led to the nationalisation of railway industries across Europe (the first was Switzerland in 1901 and the last Britain in 1948), where most of the main line railways were taken under the control of a single company so that economies of scale could be achieved. There were a number of other reasons for nationalisation, however here we will only concentrate on the economies of scale argument. This particular aspect of railway economics has become known as the traditional view (Preston, 1994), in which infrastructure and services are part of an integrated system and economies of scale in both are significant. Organisationally therefore, services and infrastructure should be part of the same (large) company.

As you may be aware however, in recent years many European countries have re-organised their railway systems, with one of the most important divisions being the separation of infrastructure from services. This is known as a vertically separated railway. Although in contradiction to the traditional view outlined above, such a division is consistent with what has become known as the revisionist view of railway economics. Under this belief, the central premise is that economies of scale are only associated with the infrastructure and not with services. Therefore, scale effects will still be taken advantage of as long as infrastructure is retained as a single entity. As regards the size of operating companies, this is unimportant as under this view there are no economies of scale associated with this activity.

Note that both the traditional and revisionist views are merely matters of opinion, and in order to get a better understanding we need to examine the empirical research that has been carried out on the topic. Examining the relevant academic literature on the topic would suggest that the traditional U-shaped cost curve (and hence S-shaped production function) apply to railway operations. For example, Preston (1994) in a study of 15 Western European (integrated) railways found diseconomies of scale for larger train systems such as (West) Germany and Britain, and economies of scale for smaller rail systems such as Ireland, Switzerland and Belgium. In an updated study, Shires and Preston (1999) found that the MES (minimum efficiency scale) for integrated railways was around about the size of the Danish and Belgian rail networks. The implication, if we only consider economies of scale, is that countries such as Britain should organise their rail system as three to four integrated railways (perhaps on a regional basis) rather than as one single national operator.

Within smaller railway systems, scale effects have been found to be even more substantial. Fillipini and Maggio's (1992) study of the Swiss 'private' rail network revealed scale effects to be considerable, leading the authors to conclude that there were considerable potential benefits to be gained from end-to-end and parallel mergers within the Swiss industry. Cowie (1999) also found scale effects to be significant in the Swiss industry. Note that Switzerland is made up of a single national operator, CFF, on the main lines, and around 60 to 70 publicly owned local railways on the local lines, varying in size from as small as 4 kilometres to up to around 400 kilometres in length. Significant scale effects in such small systems would again be consistent with a U shaped cost curve. This is because we would expect that for smaller systems the gains that could be achieved from increasing output in terms of reducing average costs would be higher than for medium sized operators (have a look at Figure 3.7 to confirm this).

Very little research however has been carried out on the effect of separating operations from infrastructure on economies of scale. Cowie (2002) however studied the British train operating companies and found that scale effects were significant, hence suggesting that the pure revisionist view, that there are no economies in operation, was not true. Size therefore, with regard to how train companies are organised, is important. Scale effects however were found to be smaller than for integrated railways, with the MES point found to be around two thirds of the output level of Preston's earlier study based upon integrated railways. The actually MES point found however, suggested that in Britain there should be around four to five train operating companies rather than the current number of 19. The research also suggested by implication that there existed significant economies of scale in the provision of infrastructure.

Case Study Questions

1. Briefly outline your understanding of the traditional and revisionist views of railway economics.
2. List what you believe to be the main sources of economies of scale in the rail industry. Once you have produced this list, indicate which arise as a result of returns to scale and which are cost savings.
3. What on the other hand do you believe are the main sources of diseconomies of scale in larger integrated railways?
4. If you were a rail industry regulator in Britain today, what other factors apart from economies of scale would you take into account when deciding on the number of operators to have in the market?

Unit Summary

In this unit the focus has been on transport costs. This began with a definition of time, in which we saw that the definition of the short run is where at least one factor of production is fixed. We then examined the behaviour of costs both in the short and long runs, and found that both the short and long run cost curves were U shaped. In the case of the former this was due to the law of diminishing returns and the latter due to economies/diseconomies of scale. We finished the unit by looking at the relationship between costs in the short and long runs. Finally, a case study of the rail industry revealed that costs in practice would appear to follow the theoretical concepts, with studies indicating that average cost curves are U shaped in both vertically and integrated railway systems.

APPENDIX TO UNIT 3: PROFIT AS A COST TO THE FIRM

This appendix to unit 3 considers the issue of profit as a cost to the firm. Maybe it is just me, but when I first came across this I found it a bit strange to consider profit as a cost, however no one really properly explained it to me and now after working it out myself it seems completely reasonable. If you likewise are in any doubt, then hopefully after reading this brief appendix the issue should become far clearer. Consider figure 3A.1.

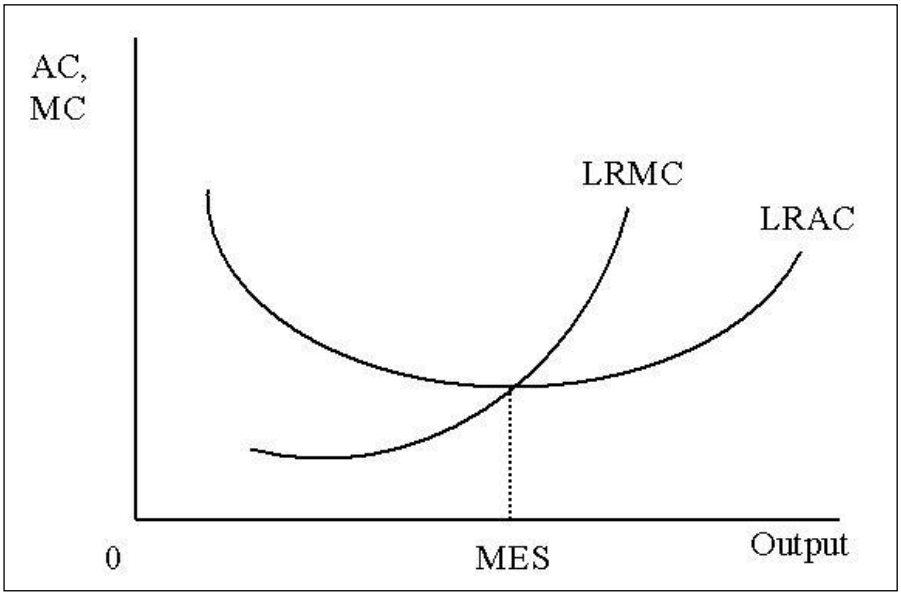


Figure 3A.1: Long run average and marginal costs

You should realise that figure 3A.1 is simply figure 3.7 from earlier. Why it is reproduced here is to concentrate thoughts on the average cost curve, in this case in the long run, however the same logic applies to the short run average cost curve. The question arises as to exactly what costs make up the average cost curve? The answer is simple, the costs of employing the factors of production in the production process. Hence it will include the labour costs and the costs associated with the physical capital, which will mainly be interest and depreciation. If that was simply it however, then it would not be showing all the costs to the firm, as the one that is missing is the venture capital costs, i.e. the costs associated with providing finance to run the business. This will normally be shareholders costs, and their reward, i.e. cost to the company, for investing in the company is paid in the form of dividends, and dividends in turn are paid out of profits. This therefore is simply a cost to the firm, in much the same manner that wages are paid to staff and all physical capital assets are depreciated (and interest paid on any borrowings required to purchase them in the first place). Profit therefore is shown in the average cost curve.

We can however distinguish between normal profit, which is shown in the average cost curve, and abnormal profit, which is not. Normal profit can best be described as 'a fair return' for investing in the firm, and is usually defined as the minimum level of profit

required to keep the factors of production in their current use in the long run. Whilst that is a bit open to interpretation as a definition, another way to view normal profit is as the opportunity cost of being in business plus some form of risk premium. In other words, what would be the return from the next best use of the funds? If this was to put the money in a bank, which in normal times would produce a steady return (except that is during the economic crisis of 2009), then normal profit should be the interest that would be gained plus some form of risk premium. Anything above that would be view as excessive, i.e. an unfair return, where the provider is gaining more benefit from the trade than the purchaser. This can be 'rectified' by reducing prices, however how this would come about is the subject of the next unit on the theory of the firm.

UNIT FOUR
THE THEORY OF THE FIRM

INTRODUCTION

In the previous units we have examined various factors relating to different aspects of transport economics. These can be broadly summarised as either concerned with the demand or supply side of the market. Under the general heading of demand, we examined issues such as demand itself, elasticity of demand and so on. Under supply, we looked at the quantity supplied to the market and the behaviour of production and costs as firm size increases. This included issues such as the law of diminishing marginal returns and economies of scale. We have also touched upon the concept of marginality and this is an idea to which we will return. What we do in this unit, titled the theory of the firm, is to put all of these concepts and ideas together to come up with a better understanding of transport markets and in particular the supply side of the industry.

Firstly we will expand upon the concept of marginality and introduce the idea of marginal revenue, before we go on and examine in some detail two particular market types, perfect competition and monopoly. We finish the unit by considering a case study of the British Bus industry, and its movement from a state owned highly regulated monopoly to a privately owned deregulated market.

Much in this unit is particularly technical, and you are advised that you may have to read through these notes two or three times before you can begin to obtain a real understanding of the issues and concepts involved.

RECOMMENDED READING

The recommended reading for this unit is:

Cowie (2010), The Economics of Transport, Chapters 6 and 7.

Mallard and Glaister (2008), Transport Economics, Chapters 5 to 8

LEARNING OUTCOMES

Once you have worked your way through this unit you should be able to:

- Understand and explain the concept of marginal revenue
- State and explain the profit maximising position of firms
- Be able to explain and illustrate the concept of perfect competition and contrast this with the other market extreme of monopoly.

- Discuss and debate the extent to which each of the UK's public transport markets corresponds with these market types and show the relevance of such market structures to the provision of transport services.

The Theory of the Firm

The theory of the firm is a very important economic concept as it underpins much of microeconomics i.e. the study of industries or individual units, rather than macro economics which is the study of the whole economy. Microeconomics as such is divided into two branches, the theory of the firm and the theory of the consumer. The former studies supply decisions by profit maximising agents to the market, whilst the latter is concerned with utility maximising behaviour from consumers given limited incomes. The theory of the firm is a neo-classical concept and has been widely used and applied to examine industry structures, efficiency and so on. It was one of the main theoretical underpinnings for Margaret Thatcher's privatisation programme in the mid to late 1980s. As such, its applications have been far and wide reaching and it has been responsible for bringing about much economic change over the last 20 years, particularly with what has followed throughout the European Union and the rest of the world in the period since Thatcher.

Despite this recent prominence, the theory of the firm has been around for a long time, coming in with the rise of neo-classical economics in the 19th Century. The whole premise of neo-classical economics is the importance of decisions at the margin, whether that be related to marginal costs or marginal revenue. We have examined most of these concepts, but we need to first introduce the concept of marginal revenue before we can go on and examine the theory of the firm.

MARGINAL COST AND MARGINAL REVENUE

As you should recall, marginal cost is defined as the "rate of increase in costs with respect to output". Another way of expressing it is as the cost of the last unit produced. Marginal revenue on the other hand is the "extra total revenue gained by selling one more unit (per time period)". In order to sell more units per time period however, the firm will have to charge a lower price than it otherwise would have done for all of its output, not just the 'extra' unit it wishes to sell. This is just simply the law of demand. You should not have the mistaken idea that if a firm sells 20 units and £5, it then sells the 21st unit at £4.50. The firm has a simple choice, it can either sell 20 units at £5 each or 21 units at £4.50 each. Which it chooses will be dependent upon a number of factors, one of course being the elasticity of demand, the other however is related to the cost of production. A stylised illustration of the marginal revenue curve therefore would look something like Figure 4.1.

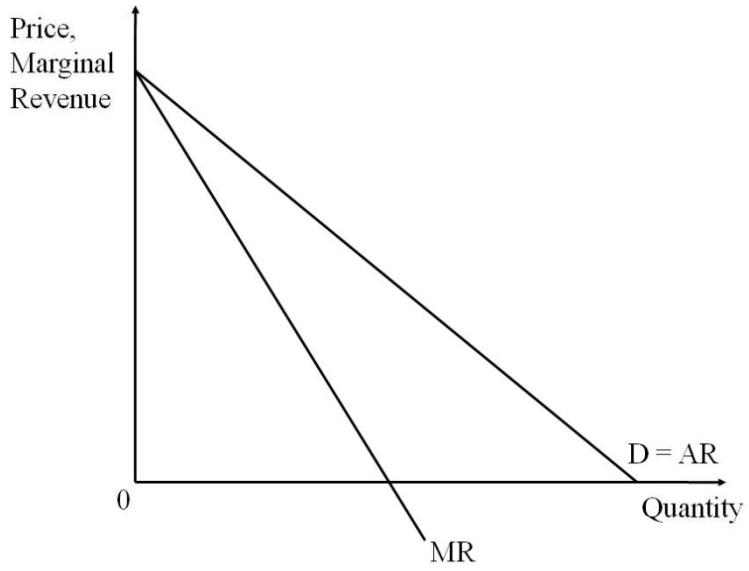


Figure 4.1: Marginal and Average Revenue Curves

The first thing to notice from figure 4.1 is that the demand curve has also been labelled 'AR', for average revenue. This should be simple enough. If we sell 100 units at £5 each, then the average revenue gained for each unit is simply the price of £5. Notice also that at all levels of output the marginal revenue is less than the average revenue, and this again is simply our marginal and average concepts – if the marginal is less than the average then it will pull the average down. Also this is due to the law of demand, as in order to sell more the firm must charge it's consumers less i.e. marginal revenue must be less than the current price, hence average revenue will fall in order to sell more. Note closer that the MR curve is twice as steep as the AR curve, and that again is a simple mathematical property.

PROFIT MAXIMISATION

We are now in a position to put all of our ideas together. Firstly however, we should re-examine the assumption of profit maximisation. This you will remember is the under lying assumption of firms when supplying the market – their aim is to make as much profit as possible. We can now formally define the firm's profit maximising position, and this is defined as:

$$MC = MR$$

Thus profits are maximised where marginal cost equals marginal revenue. In order to understand this, it is best to consider when this position does not exist. Firstly, where marginal cost is greater than marginal revenue, then this means that the last unit sold cost more to produce than the revenue which it generated. In other words, it made a loss. The firm should therefore not produce that unit, i.e. reduce production. It should continue to do so until marginal cost equals marginal revenue. If marginal revenue is greater than marginal cost, then the last unit produced actually generated more revenue than it cost to produce. The firm should therefore increase production until marginal cost equals marginal revenue, at which point profits are maximised. This profit maximising position is graphically illustrated in Figure 4.2

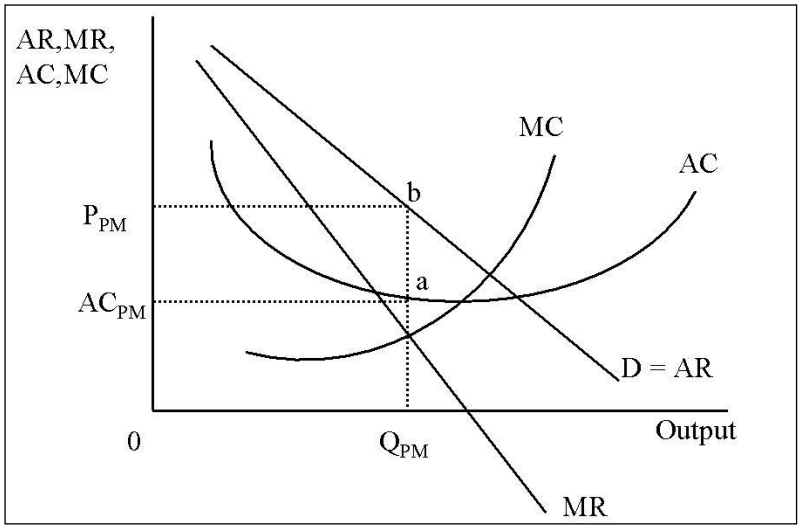


Figure 4.2: Profit maximising position for the firm

Initially it may seem that there is quite a lot on this diagram as we have made quite a substantial leap forward, hence at first sight it may seem very confusing. There is nothing here however that you have not seen before. Firstly, the average and marginal cost curves were drawn in, with the average cost curve being u-shaped as output increased, and the marginal cost curve cutting the average cost curve at its lowest point. To this was then added the average and marginal revenue curves, with the former simply being the demand curve and the latter exactly the same as outlined above.

Once this was done, the profit maximising position was easily found where the marginal cost and marginal revenue curves intersect, that is where MC = MR, which is at an output level of Q_{PM}. Once this was done, then the average cost per unit, AC_{PM}, was simply read off the average cost curve for that level of output, and the price, P_{PM}, using a similar logic with the average revenue curve (the demand curve). Total revenue is therefore given by the area outlined by $0,Q_{PM},b,P_{PM}$, total costs by the area outlined by $0,Q_{PM},a,AC_{PM}$, and profit the net difference between the two, which is the area outlined by AC_{PM},a,b,P_{PM}.

Using this general framework, we are now in a position to go on and analyse different market structures, and we begin with the two extremes of perfect competition and monopoly.

PERFECT COMPETITION

As allocative efficiency was described as the holy grail of the economics discipline, perfect competition is seen as the mechanism by which that holy grail is found. This is because this is one of the major requirements in order to achieve allocative efficiency, and we will see why exactly this is the case by examining this concept further. Perfect competition operates under a number of assumptions, of which the four main ones are:

- Freedom of entry and exit
- Homogeneous product
- High number of buyers and sellers
- Perfect information

Freedom of entry and exit means that buyers and sellers are free to enter and leave the market as they see fit, there are no obstacles preventing them from entering the market. A homogeneous product means that all firms produce identical products, thus a bus service is a bus service is a bus service. The implication is that with a homogeneous product there is perfect substitutability between the products of competing companies i.e. the demand for products is perfectly elastic. A high number of buyers and sellers means that there is no single seller or buyer that has a high degree of market power. Under a monopoly situation where there is only one seller, this firm would have a very strong market position and could exercise a very high degree of control on the market by controlling the level of output. Restricting the level of output would increase the price, while increasing output would reduce the price. The OPEC countries in the mid and late 1970s for example operated very successfully as a cartel (nee monopoly) to restrict the amount of oil on the world market and by doing so caused massive increases in the price of oil. Similarly, a small number of buyers in the market can impose great control over the sellers – grocery retailers such as Sainsburys, Morrisons and Tescos for example are in a very strong position to control their suppliers. With a high number of buyers and sellers however, this means that no single individual or organisation has any control over the market, hence all are price takers.

The last assumption is perfect knowledge. This means that buyers, sellers, potential buyers and potential sellers know everything there is to know about the market. Hence there are no trade secrets (as this may also be a barrier to entry) and all profit and market information is commonly known.

One may wish to question the highly restrictive nature of these assumptions and ask if perfect competition exists in practice? The simple answer is that the assumptions are too

restrictive to exist in reality, and perfect competition is simply a model to which reality can be compared and market failures identified. By using these assumptions however, we can draw the market and individual firm diagrams for perfect competition for both the short and long runs, and examine the market in more detail. We begin with the short run perfect competition position, and this is shown in Figure 4.3 below.

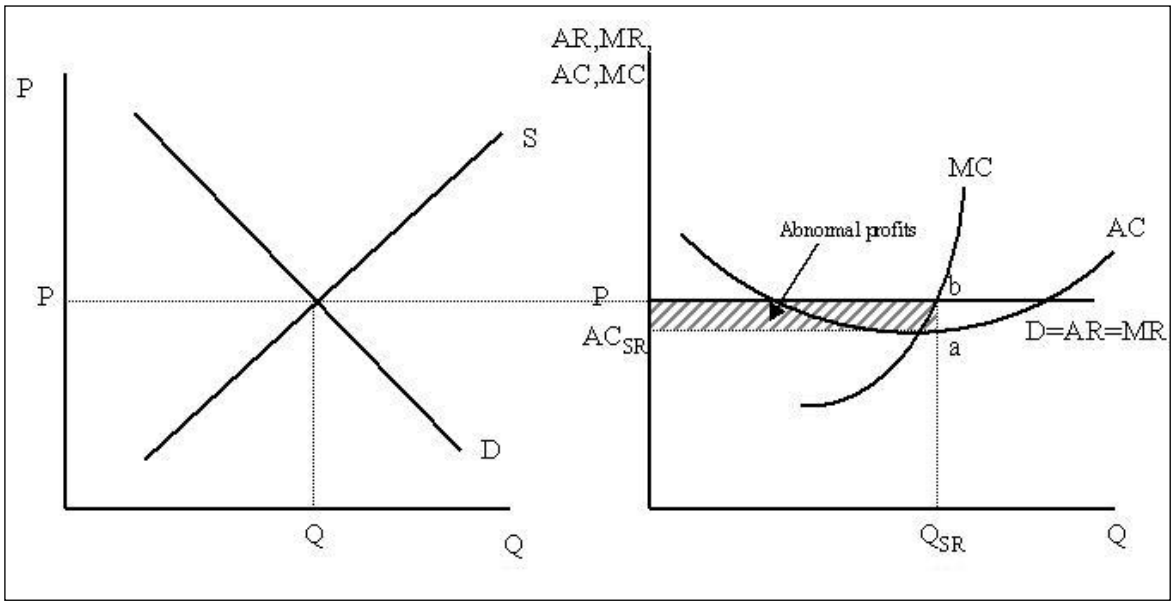


Figure 4.3: Perfect Competition, short run position (profits)

On the left is the market diagram and on the right the situation facing the individual firm. The market sets the price and due to a high number of buyers and sellers the firm is a price taker. It can therefore only sell at price P. Note that if it charges a price above P, then because goods are homogeneous and thus perfectly substitutable, all it's consumer will buy their products from elsewhere (which given perfect information they know all prices and where to purchase them from). Note also that there is little point in the firm charging a price below P, as it can sell as much as it wants at this price, hence any price reduction is simply cutting profits. This also means that marginal revenue is equal to average revenue.

On the right, the firm as normal is assumed to be a profit maximiser, hence produces at that level of output where marginal cost is equal to marginal revenue, which is highlighted by Q_{SR} for the individual firm. This is a short run position and you should be able to see that at this level the firm is making abnormal profits, as outlined by the cross hatched area AC_{SR},a,b,P. In the long run however, new firms will enter the market and compete away these profits. This is shown in Figure 4.4

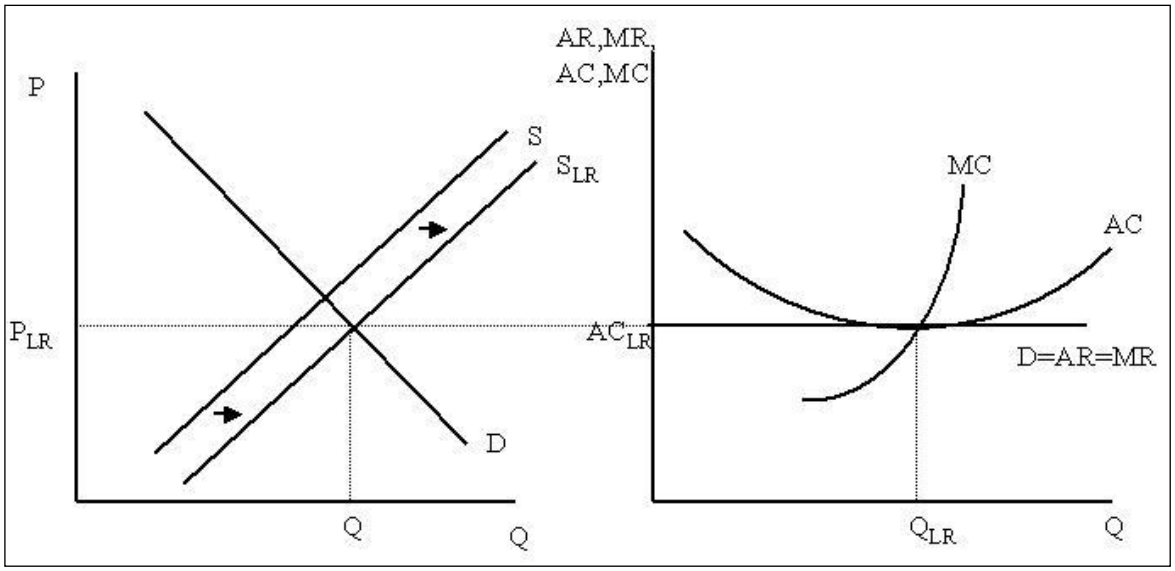


Figure 4.4: Perfect competition, the long run position

New firms enter the market as there are no barriers to entry and there is perfect information, hence they know that there are profit opportunities in this market. This shifts the supply curve to the left, hence the market price falls to P_{LR}. The individual firm adjusts its level of output to the new profit maximising position, which is at Q_{LR}. Note that at this point the firm is producing at the lowest point on the average cost curve, and as this is the long run situation, there is no incentive to change. This is why perfect competition is said to be the most economically efficient market structure, as the first two conditions of allocative efficiency, that is technical and cost efficiency, are achieved.

To further underpin these ideas, we will examine one more example - what would happen if there was long term demand shift away from a good or service. There are numerous examples of goods and services that have become obsolete, hence causing a long term decrease in demand. This is what is known as 'structural' economic change, and the most obvious example is the long term decline of shipbuilding on the Clyde Estuary. This would be represented by a shift in the demand curve to the right, as in this specific example, substitute goods in the form of lower cost non British shipbuilders take demand away from the Clyde shipyards. This would cause a reduction in demand at each and every level. This short-term position is shown in Figure 4.5

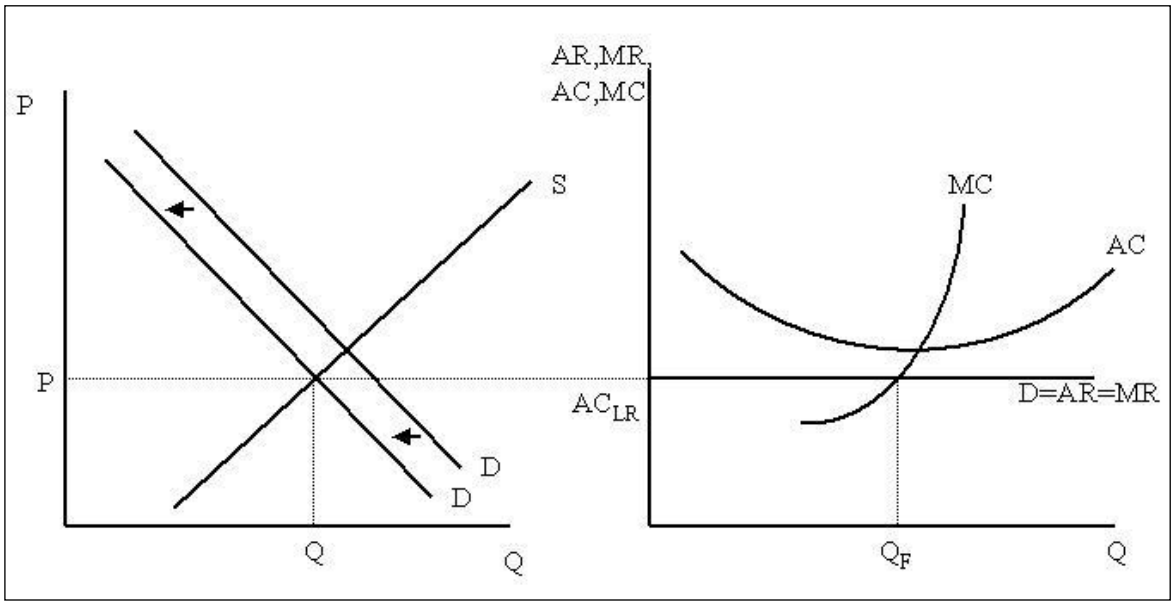


Figure 4.5: Short run perfect competition – losses

In this situation, at the market price no firm can make a profit. What subsequently occurs is a contraction in the industry (supply), which would shift the supply curve to the left as firms leave the market. This process would continue until the market was back in equilibrium at the previous long-term position. Ultimately, only those firms that produced at the lowest cost could survive in the market, but if the decline is longer term, even these firms would go out of business. They may have to move to a lower long run average cost curve, which may be done, for example, through mergers and acquisitions.

Note also from this discussion that we have what is termed consumer sovereignty. Basically firms respond to the demands of consumers, and will produce what consumers want. Firms that fail to do so either go out of business or have to establish themselves in another industry. Perfect competition therefore is said to be an ideal to be aimed at, as it does produce economic efficiency i.e. the best use of scarce resources in the long run.

A similar case may be given with the bus industry in Britain, which due to the rise of a substitute good, namely the car, has been in long term decline since 1953. This is a far more complex issue however and one that we will examine in more detail in the case study at the end of this unit.

Monopoly

At the other extreme of market structures is monopoly. A monopoly is said to occur when there is only one single supplier to the market. In practical terms however, a monopoly situation is said to occur in Britain where a firm controls more than 25% of the market. Any proposed merger therefore that would create a company that has more than 25% of the market would be investigated by the Monopolies and Mergers Commission (MMC).

Returning to theory, if a monopoly is where there is only one firm supplying the market, then the firm's individual demand and supply curves are the market's demand and supply curves, hence unlike perfect competition the monopoly profit maximising position can be shown on a single graph.

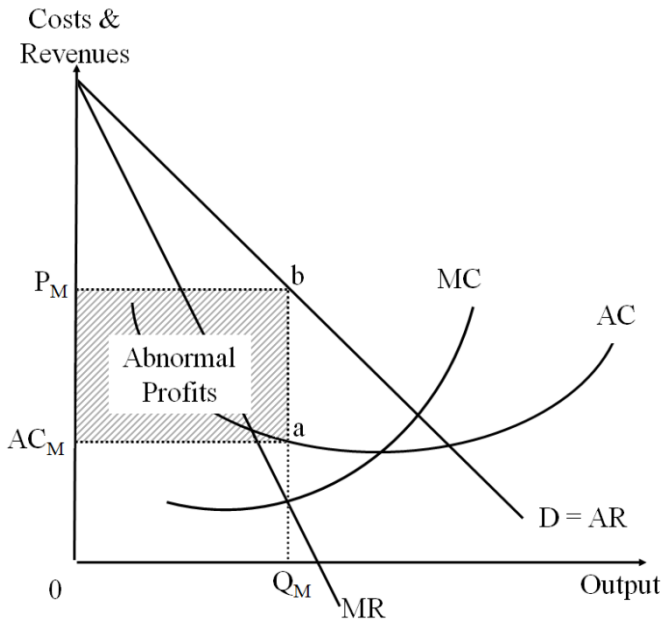


Figure 4.6: Monopolists Profit Maximising Position

Notice that this figure is almost exactly the same as Figure 4.2 above with some different labels to identify it as the monopolists profit maximising position. The monopolist faces a downward sloping demand curve (there is always an alternative to a monopolists goods or service – go without!) and marginal revenue is less than average revenue at each level of output. Extending our analysis, the firm's profit maximising level of output is found at Q_M, where marginal revenue equals marginal cost. This would produce an average cost of AC_M and the price charged would be P_M. Notice that the firm is not producing at the lowest point on the average cost curve, hence this is why monopolies are not as economically efficient as perfect competition.

At the profit maximising level of output Q_M, the monopolist would be making abnormal profits, as outlined by the shaded area AC_M, a, b, P_M. Note however that in this case, unlike in perfect competition, these abnormal profits would not be competed away in the longer run due to the existence of barriers to entry.

BARRIERS TO ENTRY

Barriers to entry are key to sustaining a short run monopoly into the longer term, as a barrier to entry blocks new firms from entering the market and competing with the incumbent. The main barriers to entry are listed below.

Firm size. Firm size is directly related to economies of scale, and this barrier to entry exists where the existing firm is of such a size that it captures all economies of scale. Its average costs therefore are at the lowest they can be. Any new firm entering the market therefore, must enter at the same size, otherwise it will be at a cost disadvantage and ultimately could be driven out of business should a price war ensure. The sheer investment involved therefore in entering at such a large scale can be a barrier to entry.

High Sunk costs. A sunk cost is a cost that cannot be re-deemed or re-claimed when the firm leaves the market. The best example of a high sunk cost, with no pun intended, is the Channel Tunnel. When this opened in 1995 this represented a very high sunk cost. If Eurotunnel were to leave the market they could not take the Channel Tunnel with them and set it up somewhere else, unlike their main rivals the ferry operators, who could leave the market by withdrawing their ferries and putting them onto other routes. A high sunk cost represents a very high risk, and as such can be a strong barrier to entry.

Legal protection. This is the most straightforward. This arises where the firm has a legal right to be the only provider/producer of a given good or service. The most common legally protected monopolies are patents, where the inventors or discoverer of a drug or process are given a legal right to be the sole producers of that good over a period of time. Others exist however. Regulatory authorities in certain (regulated) markets may nominate a particular operator to be the only operator of that service. The passenger rail industry for example, consists of 25/26 legally protected monopolies. Although there is some competition at the boundaries of areas, each train operating company has been given the sole legal right to operate passenger services in their franchised areas.

Control of the factors of production. Where one single firm controls the factors of production, then this constitutes a barrier to entry as it would be very difficult for any other firm to enter the market as it simply would not have access to any production capabilities. Although sounding a bit severe, this can relate to production processes and particularly labour skills, where one company may have all the skilled labour knowledge which is protected by covenant. The other classic example of control of the factors of production is De Beers, the world's largest producer of diamonds, which controls world supply and distribution through its London- based Central Selling Organisation.

DISADVANTAGES OF MONOPOLIES

Monopolies are normally associated with all things that are 'bad' with capitalist (market) based economies. The prime concern over monopolies are that they are anti-competitive, which results in the charging of higher prices to consumers that are faced with very little alternative other than to do without. This has led to a whole host of anti-competitive agencies, such as the Monopolies and Mergers Commission (MMC) in the UK, that have been appointed to oversee the structure of all industrial sectors of the economy. We can expand on this basic argument to look at in more depth the disadvantages of monopolies. These are outlined below.

Economically inefficient. Economic inefficiency relates to two specific issues. Firstly, as you should already be aware, unlike perfect competition where in the long run production will occur at the lowest point on the average cost curve, a monopolist will not in the long run be operating at the lowest point on the average cost curve. In simple terms this means that resources are being inefficiently used in the production of that good or service. Costs, in the form of prices of the factors of production, are under the ideal 'economic efficient' situation a signal to the producer of the best combination of these resources to use in the production process. Where these are not minimised, resources (factors of production) are not being used in their 'best' combination. Monopoly is therefore economically inefficient.

Secondly, prices charged will be higher and the output level produced will be lower than a perfectly competitive industry facing exactly the same cost curves. You should not be surprised to hear that this is best illustrated by a diagram, and this is done in figure 5.7.

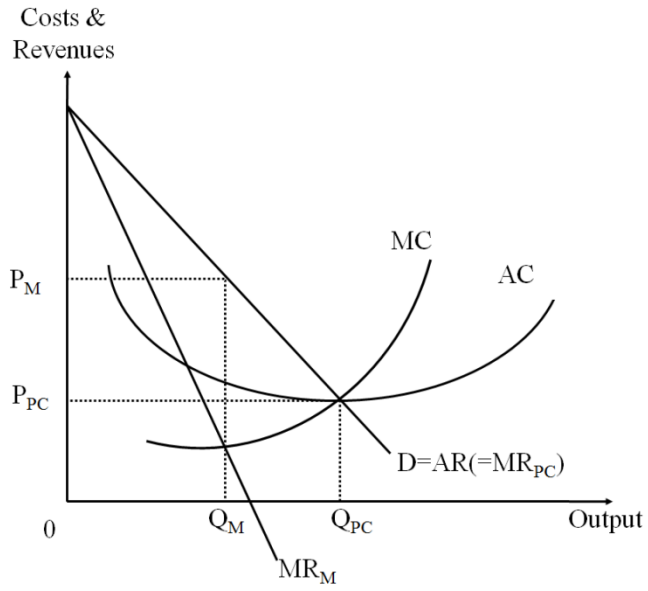


Figure 5.7: Monopoly versus Perfect Competition

Again, there is a lot on this diagram. If we begin with the monopoly position, its demand curve is the market's demand curve, which is illustrated by the demand curve D = AR (and its associated marginal revenue curve MR_M). Faced with average and marginal cost curves AC and MC, it will profit maximise where MC = MR, which is at output level Q_M and the associated price of P_M.

Turning to if this market was in perfect competition, then the average revenue curve equates to the marginal revenue curve (as all firms are price takers), hence the output level would be Q_{PC}. Under monopoly therefore, the level of supply would be less and the price charged for that quantity would be higher than if the market was in perfect competition.

Consumer surplus. This is actually another economic inefficiency argument, indeed most of our disadvantages of monopoly are concerned with the fact that it is economically inefficient. Consumer surplus is defined as that level of demand that would have been willing to pay a higher price than the market price. In a certain sense, these consumers are in 'profit' as they acquired the good or service for a lower value than the level at which they valued that good or service. Under monopoly the output level is restricted by the producer, which in effect leaves a very low level of consumer surplus. This is a debate to which we will return later, but this reason is the technical term for what is mistakenly known as a 'rip off'. This is about the balance between the utility received in the trade between the consumer and the producer, and under monopoly it is often argued that the balance is too much in favour of the producer.

X-inefficiency. X-inefficiency is a concept originally devised by Leibenstein in 1966 (Leibenstein, 1966), and lies outside 'mainstream' neo-classical economic thinking. The basis of Leibenstien's argument was that under certain conditions the average and marginal costs would be higher than they should be due to general management slack. The source of this general slackness was most prevalent under two situations. Firstly, state ownership where the lack of incentives would create such a situation. Secondly, where there was little or no competition to act as a spur to keep management control tight. Monopolies therefore, would be prime candidates to suffer from X-inefficiency.

ADVANTAGES OF MONOPOLY

Despite the disadvantages outlined above, there are a number of situations where a monopoly market structure may offer several advantages over a more competitively based system. These advantages are outlined below.

Research and Development. Where we have only one single large firm in the market it is argued that only that firm has the financial resources and clout available to it to enable it to undertake research and development. Through this research and development it can continue to make technological advances in the future that would simply not exist if production was spread around a large number of considerably smaller producers. The nub of the argument therefore, would be that monopoly can be economically efficient in the very long run. In other words, a monopolist may produce technological advances in production techniques and processes through its research and development programme that would lead to lower average costs in the very long run.

Market Size. This issue is particularly prevalent in transport markets. The basic argument is that the market is of such a (relatively small) size, that only one firm can operate in the market and achieve all of the economies of scale. If production is spread between two or three different producers, then because the major constraint is the market size, none of these producers will achieve a level of output that will capture all economies of scale. This is what would be known as a natural monopoly, where it makes economic efficiency sense that there be only one firm in the market.

Wasteful Competition. Very closely related to a natural monopoly, but the issue of wasteful competition is worth highlighting on its own. The classic textbook example of wasteful competition is where we may have two railway companies competing between two cities, as existed prior to nationalisation in 1948. Each company had it's own rolling stock and infrastructure and ran separate services. In many, if not latterly most, services however, train services ran less than half full. This is what is known as wasteful competition, as effectively double the number of scarce production resources are being used to provide a service when it would make more sense that only one company operate the service. If this was the case, most of the available capacity would be utilised. We would therefore be nearer achieving allocative efficiency.

THE THEORY OF CONTESTABILITY (BAUMOL, 1982)

William Baumol in his American Economic Review article of 1982 (Baumol, 1982) was the first to put forward the notion of the 'contestable market'. What Baumol argued was that it was unnecessary for the market to be in perfect competition in order to produce economically efficient market behaviour, what really mattered was whether the market was contestable or not. If a new entrant could enter the market and compete with the incumbent, then the threat of this potential competition would force the incumbent to act as if under a perfect (or near perfect) market structure. Rather than pursue super-normal profits therefore (i.e. profit maximise), the firm would only seek to achieve normal profits in order to deter market entry. Competitive pressures would thus be supplied by the constant threat of entry that force the firm to behave as if it was in a competitive market and hence act in an economically efficient manner. If the firm failed to do so, it would become vulnerable to entry by a lower cost operator that would eventually take the whole market and drive it out of business. This is a particularly attractive concept, particularly given that transport markets tend towards monopoly market structures, and thus the idea of the contestable market may be seen as one way in which the advantages of a monopoly can be gained without the drawbacks.

A perfectly contestable market is said to exist where entry to the market is free and exit is costless, hence no financial barriers to entry exist. It is also argued that there must be no structural barriers to the entry of firms in the long run. As noted above, structural barriers relate to industry conditions that prevent a new entrant competing with existing firms in the market. This particular point however is perhaps debatable, particularly with regard to structural barriers that relate to market conditions. For example, demand conditions in the market may act as a structural barrier to entry, as in basic terms the market simply could not support another major operator. As such, the market would not be contestable. This however appears to miss the whole point of the idea of the contestable market, as the basic premise is not that the monopolist would lose market share to a new entry but rather that the monopolist would lose the whole market to a new entrant. In other words, the market cannot support two firms, thus if truly contestable the threat is that a new entrant would take the whole market from the existing operator. On their own therefore, structural barriers would not appear to be a barrier to the contestable market. The financial risk involved in such a venture however certainly would be, thus the presence of structural market characteristics considerably increases the financial risks associated with market entry.

The same argument however would not apply to strategic barriers to entry. Where strategic barriers exist, then this does act as a clear barrier to the contestable market. Thus for example, the branding of airline services would make it far more difficult for a new entrant to gain customers from an established firm, and hence the incumbent's position would be far more secure and the contestability of the market significantly reduced.

Pulling all of these ideas and assumptions together, the market position of the perfectly contestable market firm can be illustrated graphically.

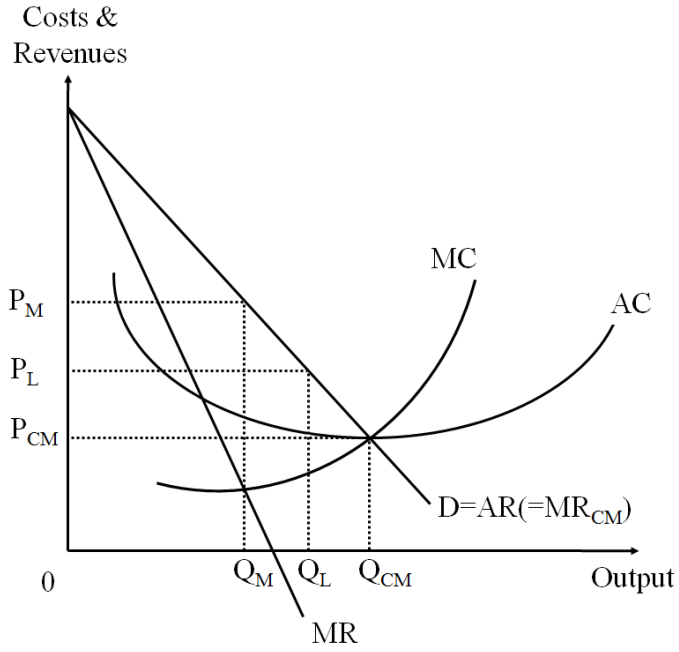


Figure 5.8: The contestable market

The basis of Figure 5.8 is Figure 5.7 above which compared perfect competition with monopoly. In this case however, if all of the assumptions of the contestable market fully apply to the monopolist then the firm will not pursue a strategy that profit maximises at output level Q_M with the associated market price of P_M. Rather, it will act like the perfectly competitive firm and set the output level at Q_{CM} with the associated price of P_{CM}. This would therefore map directly onto the position of the perfectly competitive firm illustrated earlier.

In simple terms however, very few truly contestable markets exist in practice as the assumptions of the contestable market are highly restrictive, particularly those relating to zero entry and exit barriers. Like perfect competition, the assumptions are virtually impossible to find in reality. Low cost airlines are said to be near such a model of competition, as aircraft can be leased on entry and returned to the leasing company on exit, hence significantly reducing the entry and exit costs. Whilst that may overcome the major financial entry barrier, others still remain with regard to the availability and attractiveness of airport landing slots and various other factors concerning the infrastructure. Even in such a market therefore, there will always be a cost associated with market entry and exit[1]. In reality therefore, a monopolist in a potentially contestable market will set its level of output somewhere between Q_M and Q_{CM} dependent upon the level of barriers to entry. In order to illustrate this idea, a Q_L level of output has been

added to Figure 5.7. If entry and exit barriers are relatively low, then this would indicate that the market is highly contestable and the output level Q_L will tend towards the perfectly competitive market position of Q_{CM}. If entry barriers are relatively high however, then the contestability of the market is severely compromised and the firm's output level will tend towards the monopolist position of Q_M. This is Bain's idea of a limit price (Bain, 1956), where the firm will not set a price that maximises profits but rather set the price at a level that will deter market entry. The extent to which the limit price P_L deviates from the perfectly competitive market price of P_{CM} will be directly related to the scale of entry and exit barriers. In most situations however, some degree of abnormal profits will still be achieved and maintained in the long run due to such barriers.

Self-assessment exercises

1. Without referring to the text, can you draw a short run perfect competition model showing abnormal profits being made? Can you then draw the long run position and explain why under perfect competition the industry would move from the short run situation to the long run position?

2. On blank sheet of paper, draw a (longish!) straight line. On the left side of the line, write the word 'Monopoly' and on the right side of the line write the words 'Perfect Competition'. Now consider the following transport industries/markets:

   i. The Glasgow Bus Market
   ii. The Scottish Rail Market
   iii. The Road Haulage Industry
   iv. The Cross Channel Market
   v. British Domestic Air Services
   vi. The British Rail Freight Market
   vii. Edinburgh Taxis
   viii. Scottish Interurban bus services
   ix. The British Public Transport Market

   Now place where you think each of these industries/markets should fit on your chart between Monopoly and Perfect Competition. You should give a rationale for each of your positionings.

   What have we learned (if anything??) from this exercise?

Case Study – The British Bus Industry

The British Bus Industry provides an interesting case study of changing market structures, ownership issues and regulation. Until the industry reforms of the early 1980s, the industry for the previous fifty years had been largely governed by the rules and regulations laid down by the Transport Act 1930. That Act regulated all bus services, with the country being divided into 11 different regions (subsequently reduced to 9, then 8 and now 7) for regulatory purposes. A Traffic Commissioner had complete control over all bus services in each of these regions. They regulated frequencies on routes, set prices and strictly controlled entry and exit to the market as a whole and on each specific route.

The industry supply side structure from 1968 onwards was effectively a monopoly, or more precisely, a series of co-existing local monopolies. All major bus companies were state owned, with very few private operators on staged services. Two national operators existed, the National Bus Company (NBC) in England and Wales and the Scottish Bus Group (SBG) in Scotland. There was a further 50 to 60 bus companies that were local authority owned. These companies had exclusive rights to operate in their respective towns and cities, while the NBC and SBG had exclusive rights to operate everywhere else. There was very little direct competition.

If we take Glasgow as an example, city bus services were operated exclusively by the Glasgow Corporation, whilst the surrounding outlying areas in counties such as Lanarkshire and Ayrshire were served by subsidiaries of the SBG. There was some limited competition on routes from the outlying areas to the city centre, where both the Corporation and SBG would operate services along the main corridors. Prices and frequencies however were strictly controlled, thus the two main mechanisms by which companies could effectively ‘compete’ were nullified.

With the coming to power of the Conservatives under Margaret Thatcher in 1979, this situation was about to dramatically change. Thatcher and her deputies (notably Keith Joseph, Minister of Trade and Industry), encompassed the economic beliefs of Milton Friedman and what become known more generally as ‘New Right’ ideologies. At the very heart of this ideology is the belief in the power of the market to find the optimal solution to produce economic efficiency, equity and general well being. An early piece of Thatcher legislation was the Transport Act 1980. This amongst other things removed all economic regulation from the long distance bus market, or ‘non-staged’ services (i.e. those above 25 miles in length). Any operator was free to operate on these routes and charge whatever price they wished. In other words, this had been moved from the previous situation of a monopoly to one of increased competition. This Act in itself was wholly successful, with increased patronage, reduced prices and a significantly increased network of cross country bus services. Interestingly, this also gave Stagecoach a start in the bus industry.

Following the success of this act, the Transport Act 1985 took things further, a lot further. The main problems perceived with the bus industry at that time were falling patronage numbers (these had decreased every year since 1953), increasing costs and increasing levels of subsidy. In order to attempt to reverse these trends, the 1985 Act removed all economic regulation (i.e. price and capacity constraints) from staged services throughout the country except in London, which would remain as a regulated market to be dealt with at a later point in time. The 1985 Act also re-organised the supply side of the market, resulting in the creation of around 150 new bus companies. The NBC was divided into 72 regionally based operators and the SBG into nine. These were sold to the private sector in the form of a share issue (for National Express) and private sales, either to the existing management of the company or to existing bus companies such as Stagecoach. Sales were restricted however, with no single purchaser being able to buy more than 3 NBC subsidiaries and 2 SBG subsidiaries. NBC was sold off between 1986 and 1988 and SBG over the period 1990/91.

The 1985 Act required Local Authorities to organisationally separate their bus operations from the authority. Prior to that time in many instances the bus operations had been like any other department in the authority. These were to be set up as 'arms length' companies, which in simple terms means completely independent of the authority. They were to no longer receive direct subsidy, as most had run at a loss with the authority making up the net difference at the end of the year. These measures were seen as a prelude to privatisation, and companies were sold off from 1986 onwards, with the last being sold in 1995.

The final measure of the 1985 Act was to completely de-regulate the bus market outside of London, with no restraints on capacity or on the setting of fares. The only requirement was a purely administrative one in that new routes or withdrawal of services had to be registered with the Traffic Commissioners for a notice period of 42 days. Local authority powers in the planning and control of bus routes were severely limited, however they could still specify services on the grounds of a social necessity and put this out to competitive tender, where bus companies would lodge bids to run the service for a given amount of subsidy.

Finally, unlike other privatisations, there was to be no industry regulator. When British Telecom had been privatised, OFTEL had been set up to regulate the industry, as was true of OFWAT for water and OFGAS for gas. It was felt however that the existing regulator competition agencies, in the form of the Monopolies and Mergers Commission and the Office for Fair Trading, would be able to handle all regulatory issues arising from the industry.

The rationale of this whole industry reform was based on the premise of the power of competition. The 'vision' as such is that the bus industry would be made up of a large number of small to medium sized operators. Many of these would be owned by former managers of state owned companies, and these companies would compete with each other in order to better meet the needs of the passenger in terms of price, quality and

overall standard of service. Through this process the first of the aims of privatisation, to reverse falling patronage, would be achieved. The second, to reduce costs, would be achieved through the power of the private sector to better control costs, and thus the combination of increased revenue through increased patronage and decreased costs would eradicate the need for subsidy, thereby achieving the third aim.

What Happened After 1985?

All bus services outside of London were de-regulated on 23rd October 1986. This caused massive disruption in most major cities, with Glasgow suffering particularly badly as the streets were swamped by buses of various vintages. There then followed the period of what became known as 'the bus wars', where some cut-throat competition took place in order to expand company market shares. One of the worst examples of this occurred in Darlington, where the local authority sought to sell the authority owned bus company to the private sector. After a short bidding period, which included bids from Stagecoach, the authority gave preferred bidder status to a company called Yorkshire Traction, a private limited company with a fleet size of around 2000 buses. Following this announcement, Busways, a Stagecoach subsidiary, re-entered the Darlington market with a vengeance, offering free bus services and enticing drivers away from the local authority bus company to work for Stagecoach for a £1000 single payment. The local authority company could not be sustained under such severe competitive pressures and went bust, leaving the Darlington market as a Stagecoach stronghold.

This is just one example, but the bus wars period included constant changes in timetables, the duplication of well served routes and the withdraw from more rural services, dangerous driving manners and some dubious business practices. Although significantly reducing costs and subsidy, the privatisation/de-regulation measures failed to stem the decline of passenger numbers, and with rising prices and the creation of considerable confusion of bus services, even led to increasing the downward spiral. This unhappy period for the bus industry came to an end with the emergence, through acquisition, of three major operators, Firstbus, Stagecoach and Arriva. These bus companies now dominate the market, with around 50% of passenger revenue between the three of them. What has emerged is now a series of territories, where one bus company dominates in one area with limited competition from another. In Scotland for example, the rough break down is as follows:

Location	*Dominant Operator*	*Competition*
Ayrshire	Arriva	Stagecoach
Glasgow	Firstbus	Stagecoach
Edinburgh	Lothian Buses	Firstbus
Fife:	Stagecoach	Firstbus

Dundee:	National Express	Yorkshire Traction
Aberdeen:	Firstbus	Stagecoach

This is only a rough sketch of bus territories. Ayrshire for example, is roughly split between Stagecoach and Arriva, and in Glasgow Firstbus dominants with competition from Stagecoach along the main corridors. In Edinburgh, Lothian Buses have around 85% of the Edinburgh market, with Firstbus having most of the remaining 15%. Whilst the above illustrates the case in Scotland, the pattern of a dominant operator with some limited competition from one of the major operators is virtually repeated throughout the whole of Britain with the possible only exception of Oxford, where a truly competitive market still exists.

In many cases these markets constitute local monopolies and certainly are quite far away from a highly competitive market. If you remember from earlier in this unit, this need not necessarily be a 'bad' thing as long as the market is contestable. These markets however are neither competitive nor contestable. If for example Stagecoach was to enter the Glasgow market on a much larger scale, Firstbus would be likely to retaliate with entry into Stagecoach strongholds. Such a threat of retaliatory action is therefore a very strong barrier to entry.

Whilst painting a very negative picture of the bus industry, it is not all bad news. Whilst companies were expanding through merger and acquisition, company growth as such could be gained relatively easily. Now however with opportunities for growth in firm size through such measures severely limited, companies have to examine other methods of maintaining and increasing market share i.e. more focus on the consumer. Some success has been achieved with bus quality partnerships, where the local authority agrees to put in bus priority measures such as bus lanes and generally improve the infrastructure (bus stops, bus stations, passenger information etc). The company in return agrees to invest in new buses and provide an overall higher quality of service. The second positive issue to arise is increased investment. At the time of privatisation the industry was heavily characterised by aged bus fleets, with average fleet ages being in the order of 10 to 12 years. Since the mid 1990s however, investment levels have risen considerably and now average fleet ages are of the order of 5 to 7 years. It is also worth stressing that higher prices need not necessarily be a 'bad' thing if they allow operators to fund investments out of profits. The final piece of positive news is that in year 2006/7 for the first time since 1953, patronage levels actually rose for bus transport outside of London.

Case Study Questions

1. To what extent do you think that the 'vision' that was foreseen through the 1985 Transport Act meets the conditions of perfect competition?

2. One of the main reasons cited for the growth in bus company size after 1985 was due to economies of scale – what are the main economies of scale in bus operation and do you believe these to be significant?

3. What do you see as the main barriers to entry to the bus market?

4. Do you believe that any of the advantages of monopoly apply in this case study?

5. What does this case study tell us about government intervention in the structure of transport industries?

UNIT SUMMARY

In this unit, we have considered the theory of the firm. The natural starting point was a summary of the public transport product and this was followed by a description of the twin economic concepts of profit and marginality. We then examined the idea of perfect competition. This is an important concept in many respects, because that is how transport markets 'should' work. If in reality we had perfect competition, then we would have economic efficiency and social welfare would be maximised. This is an important point, as it shows what should happen, but the reality in most cases is very different. It therefore allows us to identifying where things are going in wrong, i.e. a breach of one or more of the assumptions of perfect competition, and this constitutes market failure, i.e. the market doesn't work. It therefore allows us to focus on what should be done in order to put this right. We also examined monopoly, which actually is an aspect of market failure, as it breeches several of the conditions of perfect competition. With transport markets we tend to find in reality that most tend towards this or some other form of imperfectly competitive markets. We continue with the theory of the firm in the next unit, looking at pricing under different market conditions, before we return in the last two units as to the options available to transport and regulatory authorities to intervene in transport markets to steer the market towards some form of perfectly competitive or contestable market.

APPENDIX TO UNIT 4: THE MARGINAL COST CURVE AS THE SUPPLY CURVE

Consider the following diagram which shows the market position of the perfectly competitive firm as demand increases over time:

Figure

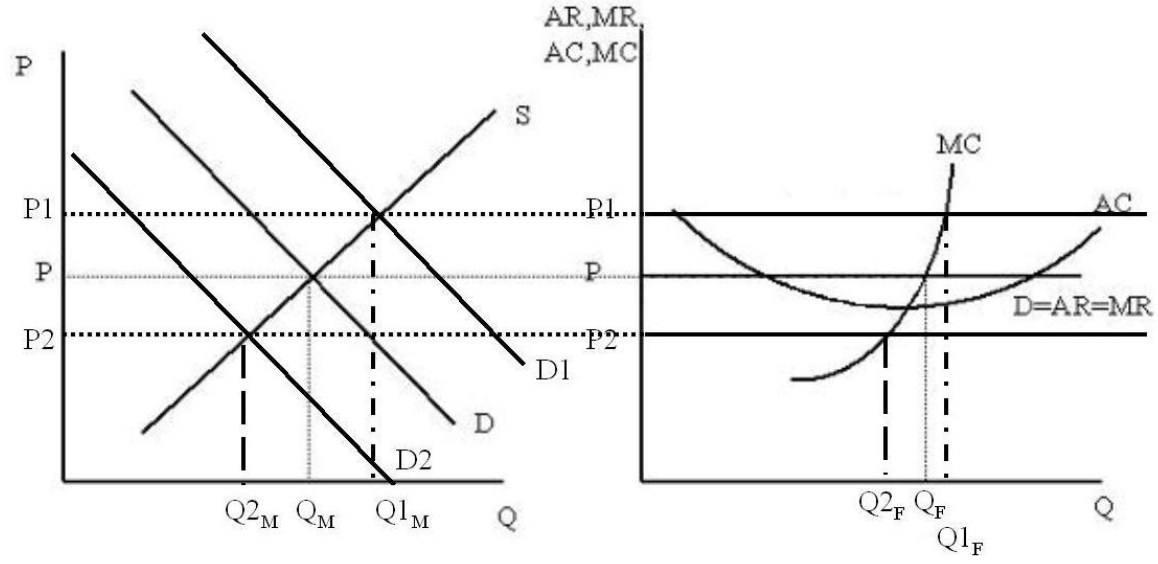


Figure 4A.1: The perfectly competitive firm's level of output under three different demand conditions.

What this shows is the individual firm's output level as demand is changing. Hence when the demand curve is at D2, the individual firm is producing where MC = MR, which is a quantity of $Q2_F$. When demand increases to D, again the individual firm will produce where MC = MR, this time that produces an output level of Q_F. Finally, when the demand curve shifts to D1, the firm again produces where MC = MR, this time at output level $Q1_F$. When viewed together, you should notice that the firm is in fact producing along its marginal cost curve. The marginal cost curve is therefore the individual firm's supply curve. This is more clearly shown in figure 4A.2.

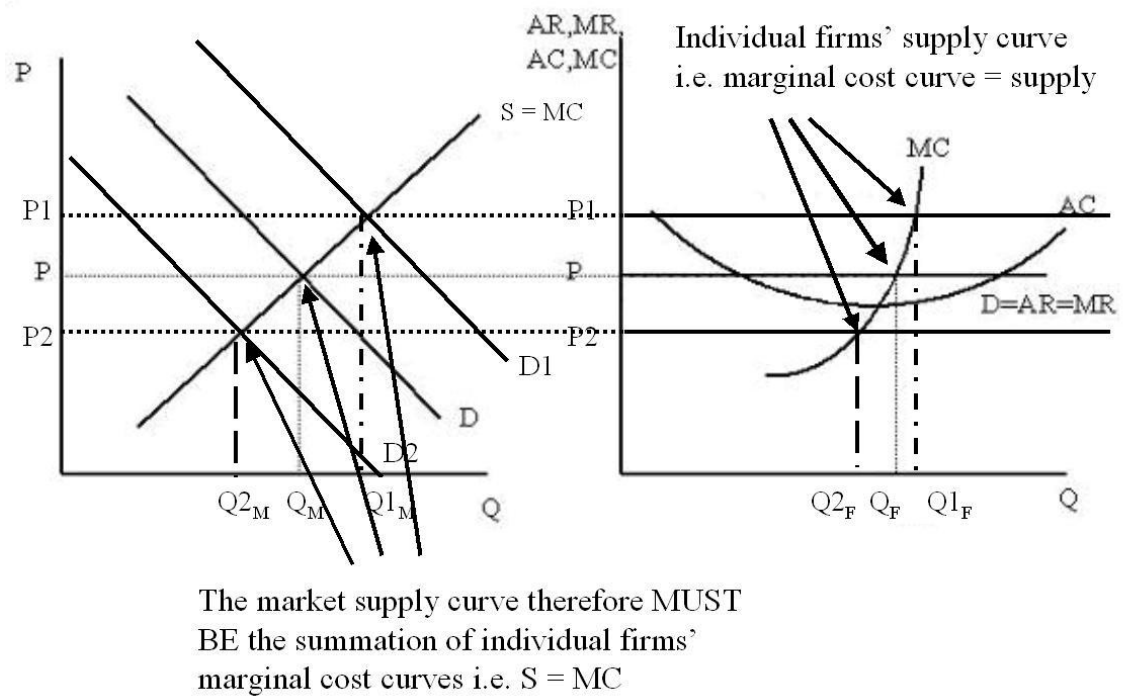


Figure 4A.2: the supply curve as a summation of individual firms' marginal cost curves.

As illustrated on the figure, the market supply curve must therefore be the summation of individual firms' marginal cost curves. As a result, the supply curve can be labelled as S = MC.

In order to complete the picture, although not shown by our graphs relating to perfect competition, the market demand curve can be labelled as D = MB. This is because the demand curve measures (in financial terms by how much individual's value something) the marginal benefit gained by consuming one extra product or services. This is one of the reasons, for example, why it slopes downwards from left to right, due to the law of diminishing marginal benefit, i.e. as consumers consume more of something, they receive less benefit from each subsequent unit consumed, and hence being rational(!), would be willing to pay less for it.

An understanding of these two ideas is useful, if not vital, for the remaining units of these course notes.

UNIT FOUR B
THE PRICING OF TRANSPORT SERVICES

INTRODUCTION

In the previous unit we looked at costs and market structures incurred by public transport operators under the two main headings of perfect competition and monopoly. This unit is concerned with another aspect of the operations of the public transport firm, namely their pricing policy. This builds upon the last unit as we examine pricing policy with regard to different market structures, and introduces a third structure known as the oligopoly market. The unit ends with a detailed look at the process of yield management. Essentially, yield management concerns the ability of management to maximise revenue from a pre-defined activity (e.g. a scheduled flight or train journey) by a combination of price discrimination and product differentiation. These two concepts are particularly important in medium and long haul transport markets, both for passengers and for freight. Note that this unit is included for supplementary reading.

We have assumed up to now that firms seek to maximise profits. As has been noted earlier however, this may be only in the long run, and profit maximisation may be pursued in the short run through other objectives. The principle objective of a pricing policy of a transport operator for example is to maximise revenue. This can be done in two ways:

- Extend the market size; and
- attract customers from other operators, thereby increasing market share.

It should be noted that revenue will come from sources others than tickets sold, such as advertising on public transport. These sources of income will also be maximised to their full potential by the operator to make as much income as possible. The unit mainly concentrates however on the pricing of the primary product, that being the transport service.

RECOMMENDED READING

The recommend reading for this unit is:

Ison, S. (2010), The Pricing of Transport Services, chapter contributed to Cowie (2010), The Economics of Transport.

Cole S (2005), *Applied Transport Economics*, Chapter 5. Pricing policy.

Button K J (1993), *Transport Economics*.

Hibbs, J (2003), *Transport Economics and Policy*, Chapter 3 – The Importance of Pricing

LEARNING OUTCOMES

Once you have worked your way through this unit you should be able to:

- Understand the different market conditions that affect pricing
- Outline the model of oligopoly, highlighting its underlying assumptions
- Grasp the economic concepts of yield management and market segmentation;
- Appreciate the role of price discrimination and product differentiation in the process of yield management;
- Apply the concepts of price discrimination and product differentiation to public transport examples; and
- Understand the range and type of tickets available on public transport.

Pricing under Different Market Conditions

How a firm sets its price will be dependent upon a number of factors. Probably the most important dimension of these however is the number of competitors that the firm is facing in the market place. As you should be aware, this dimension is best characterised by the different market structures that we examined in the last unit. We thus begin this unit by examining pricing policy under different market conditions. Two of these structures, perfect competition and monopoly, we have studied before, however we will also introduce a third market structure, known as oligopoly. We begin with perfect competition however.

PERFECT COMPETITION

Of all the market structures, pricing under perfect competition is the most straightforward. From the key assumptions that underpin the model of perfect competition, the most important one with regard to pricing is that there are a large number of buyers and sellers. All firms consequently are price takers i.e. the market sets the price. This is shown in figure 4b.1.

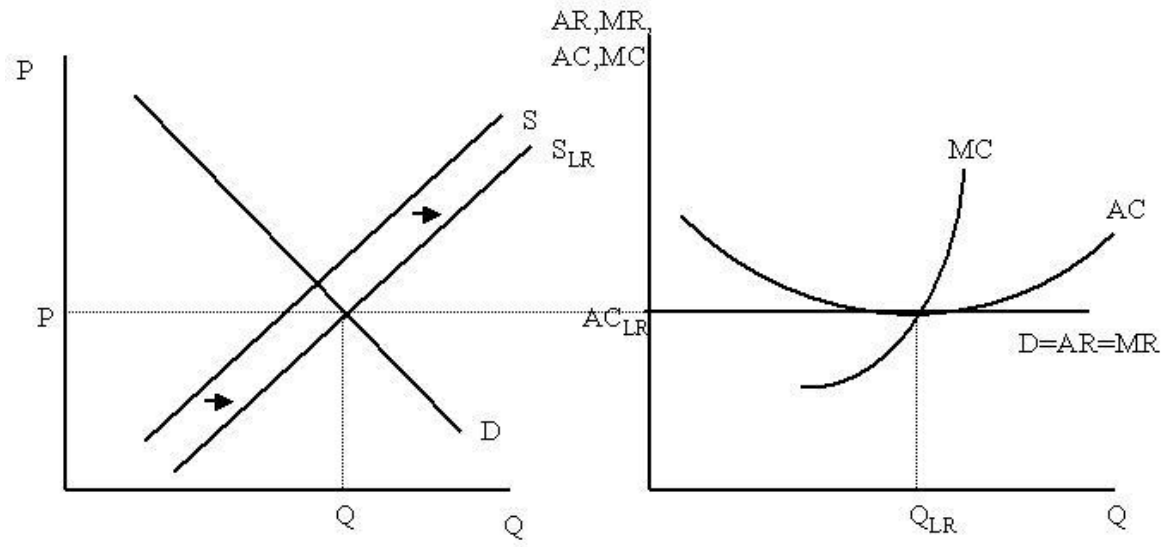


Figure 4b.1: Pricing under Perfect Competition

We have seen this diagram before in the last unit. Under perfect competition as all firms are price takers, they in effect do not have a pricing policy under this market structure as the market sets the price. If the firm chose to charge a price above the market value, all consumers would switch to other firms, as we have perfect substitutability (homogenous good or zero product differentiation) and perfect information (e.g. knowledge of prices and products). If firms on the other hand chose to charge a price below the market level, this would only be cutting profits as the firm can sell as much as it wishes at the market price.

Situations like those conditions pertaining to perfect competition tend to occur where there exists a highly competitive market. Under these conditions firms must charge the lowest price possible (and hence have the lowest costs possible) otherwise lower cost rivals will drive the firm out of business. As highlighted in Unit 4 however, perfect competition does not exist in practice, but rather is an ideal at which to aim and use to analysis real world situations. Furthermore, in transport markets now there is also little evidence of any markets that even closely resemble perfect competition. The only possible market with regard to transport industries is with regard to Road Haulage, where the market is made up of a high number of small operators. Public transport markets however more closely resemble an oligopoly market structure.

PRICING UNDER OLIGOPOLY

We have examined perfect competition and monopoly in the last unit, however here we introduce the idea of an oligopoly. Whereas the first two market structures were quite straightforward, oligopoly tends to be one of the more messy pieces of economic theory. The important difference is that where as there is only one type of perfect competition 'model' and one type of monopoly 'model', there exist different theories of oligopoly. We

will simply cover the basic general ideas surrounding oligopoly in this section, and particularly those aspects pertinent to transport markets.

Oligopoly lies somewhere between perfect competition and monopoly if assessed on the basis of the scale of competitiveness in the market. It is undoubtedly however closer to monopoly than perfect competition, as the number of competitors tends to be small in number. Virtually all public transport markets would probably be classified as an oligopoly. The underlying general assumptions of oligopoly are:

1. Few sellers, many buyers

In an oligopoly market, there exist few sellers and many buyers. The main implication from this assumption is that when making market conduct decisions such as pricing decisions, firms will take into account rivals likely reactions to their market conduct. This contrasts both with perfect competition where there are so many other sellers in the market place the firm cannot take such actions into account, and monopoly where there are simply no other firms in the industry to take into account.

2. Barriers to entry are significant

With significant barriers to entry such as those reviewed in the last unit, firms within the industry have a degree of 'protection' from new entrants. When considering pricing decisions therefore, little account needs to be taken of any potential competition.

3. Non price competition

This is one of the key features of oligopoly. Where there are very few sellers, price wars tend to damage all firms in the industry and benefit none. What tends to happen under oligopoly therefore is that firms compete on factors other than price, hence all charge the same price. These other aspects include issues such as frequency of service, market position, flexibility of tickets, special offers and so on. We will return to this aspect of oligopoly later in this unit.

4. Product differentiation

Unlike perfect competition, where all firms sell the same product, or monopoly, where only one product is sold, under oligopoly there exists a degree of product differentiation. The main consequence of product differentiation is that firms in their market conduct decisions can expect a degree of brand loyalty to exist. Hence if the firm was to charge a price above their main rivals, they can expect to retain some, if not most, of their customers.

This aspect of oligopoly however is where the theory begins to be a bit 'messy'. This is because in many oligopoly markets there is in reality no differences in the products that are marketed i.e. there is no product differentiation! The classic example is the petrol market, where what you put in your tank is absolutely identical stuff whether it is Shell, Esso, BP, Morrisons, Sainsbury's or whatever. Bus and airline journeys are exactly the same, where in basic terms the basic product is identical. Coke (the drinking kind!) is exactly the same – it's black, fizzy and bad for your teeth! These however are all examples of oligopoly markets. What is important therefore is the perception of product differentiation, and hence in turn what becomes important under oligopoly is advertising. Through advertising, firms build up a perception of differences in competing products and thereby build up consumer loyalty. This gives them a degree of power over the market and hence more flexibility in setting the price.

5. Tacit Collusion

Oligopolistic industries are said to experience what is known as tacit collusion. Tacit collusion means there is a hidden degree of co-operation. This does not mean hidden from regulatory authorities etc, but rather that under such a market structure there is a strong incentive for firms, to a certain extent, to co-operate rather than compete with each other. Under oligopoly, in an ideal situation firms should fully co-operate and take decisions as a single group of companies. This is what would be known as a cartel. Cartels enable oligopoly firms to act like a monopolist and hence maximise profits of all. Such actions however are anti-competitive and illegal under European Union competition laws. Tacit collusion however does not refer to such illegal 'active' co-operation but rather to unspoken/'passive' co-operation where it makes sense for firms to relax competitive pressures against each other in certain situations.

Using these assumptions, we can construct the oligopoly market position and thereby determine pricing policy under this market structure. This is done in Figure 4b.2.

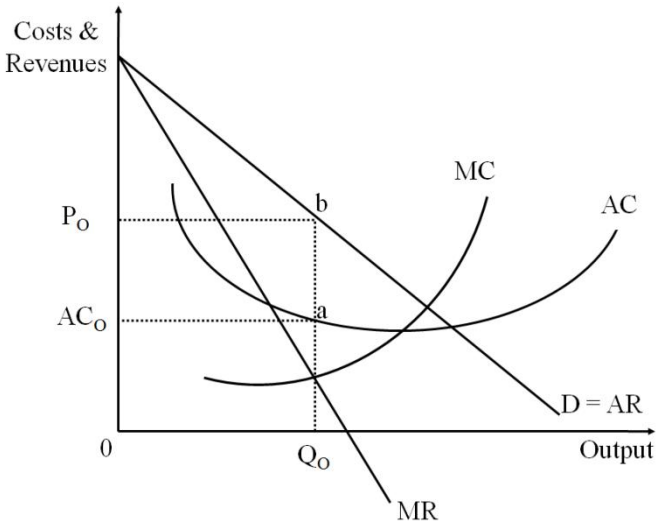


Figure 4b.2: Pricing under Oligopoly

Again, this is simply a variation of a diagram that we have seen before, as this is the basic profit maximising position of the firm. Under oligopoly, as there is a degree of consumer loyalty the firm faces a downward sloping demand curve from left to right – if it increases its price it will lose some but not all of its consumers. In order to sell more units in any given time period, the firm must sell all products are a lower price, hence at each level of output MR < AR (although we will relax this assumption later in this unit). Costs behave as before, with average cost first falling due to economies of scale, are minimised at the Minimum Efficiency Scale point, and then rise due to diseconomies of scale. Finally, the marginal cost curve cuts the average cost curve at its lowest point.

Our overriding assumption of firms behaving as profit maximisers still applies, hence the firm produces at that level where MC = MR. In the above diagram, this is at output level Q_O, which gives an average cost of AC_O and a price of P_O. Note that at this level of output the firm is making abnormal profits of AC_O, a, b, P_O, and also that the firm is not producing at the lowest point on the average cost curve as AC_O is above the MES point. It is therefore not economically efficient. You should realise that this diagram is identical to that of the profit maximising position of the monopolist, the only difference is the labelling of the relevant points.

This is the basic neo-classical position, in which the firm sets the output level at the profit maximising position, and then charges a price that the market will bear (as shown by the demand curve). Note however that the firm influences the market primarily through its output level supplied to the market, and that by restricting output it can increase price. For a bit of a less 'black box' type approach, other theories concerning pricing policy with regard to oligopolistic markets have been developed, the most notable being the issue of price leadership.

PRICE LEADERSHIP AND THE KINKED DEMAND CURVE

An important and unique aspect of the three market types we have examined, is that under oligopoly when setting its price the firm will take into account the actions of other companies when doing so. This has led to the development of the Kinked Demand curve, which is illustrated in Figure 4b.3.

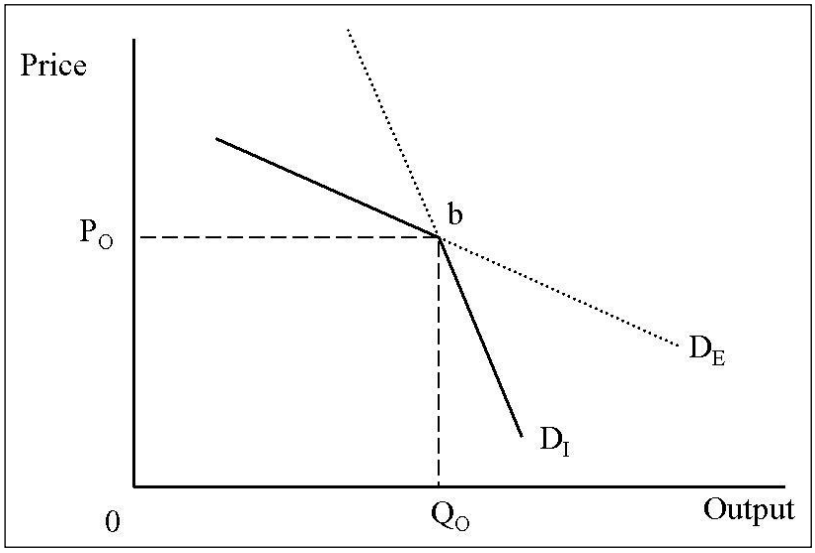


Figure 4b.3: The Kinked Demand Curve

As can be seen from Figure 4b.3, the Kinked demand curve is actually made up of two different demand curves. These in turn represent different reactions from rivals to a firm's change in price. The demand curve D_E represents a more elastic demand, whilst D_I is more inelastic. The basic theory is that if a firm was to increase its price, demand would follow the path of D_E. This is because it is assumed that no other firm will increase price, hence it will experience a substantial drop in quantity demanded. It will retain some of its consumers due to brand loyalty, however many will switch to what are now less expensive rival products. As demand is elastic, then any increase in price will actually lead to a decrease in total revenue, hence there is little incentive to take such action.

If the firm was to cut its price however, the theory is that other firms would follow suit for fear of experiencing a massive fall in demand, as would happen if demand followed D_E. Demand would therefore follow the demand curve D_I. Although the lower price will encourage more consumers to purchase the product, because other firms have followed suit, these 'new' consumers will be the only increase in the quantity demand for the individual firm's product. The firm's demand curve is therefore kinked at the market price of P_O.

The theory of the kinked demand curve helps to explain why firms in an oligopoly market charge the same price and also helps to explain why there is little, if any, incentive to compete on the basis of price. What it does not do however is explain how the price came to be at the level that it is in the first place. In order to examine this issue, we need to consider price leadership.

PRICE LEADERSHIP

The idea of price leadership is very simple. The basic idea is that there is one dominant firm in the market. When setting its price, it will take into account the actions of other firms, but likewise other firms take the dominant firm's price as given when setting their own price. This obviously is dependent upon tacit collusion. Once the price is set, due to the kinked demand curve, there is little incentive for any firm to change its price.

Pricing Under Monopoly

The one key difference between pricing under monopoly and pricing under the two market structures that we have already examined is that under monopoly there are no other competitors. Furthermore, barriers to entry prevent new entrants from entering the market in the future. Consequently, when the firm sets the price, it does not have to take into account either actual or potential competition. Note however that under monopoly consumers still have an alternative, and that is to do without. There is therefore a limit on the maximum price that can be charged.

The main consequence of the above however is that under a monopoly market structure the firm has a number of different options when setting price. It also has the highest degree of flexibility. Under perfect competition for example, all firms were price takers hence had no input into the process of setting the price. In oligopolistic markets the firm had two choices – either compete on price or compete on other issues, and we saw that there were very strong incentives for the former. Under monopoly however, the firm has a number of different options with regard to setting price. The first is the most basic, and this maintains the assumption that all firms are profit maximisers. Hence the profit maximising output level is found where marginal cost equals marginal revenue. This position is shown in figure 4b.4.

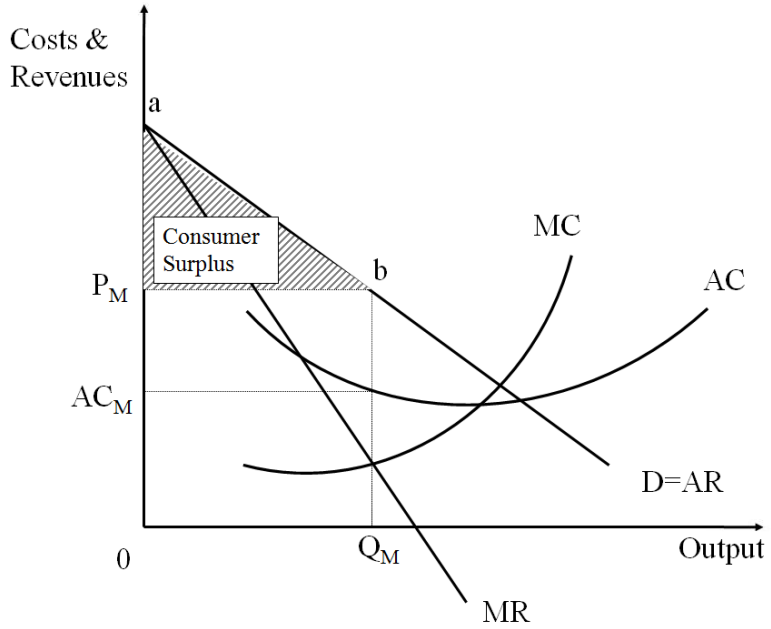


Figure 4b.4: Pricing under monopoly

As before, this is a variation on a diagram we have seen before. For the monopolist, the firm faces a downward sloping demand curve, hence marginal revenue is always less than average revenue at each level of output. The profit maximising output level is found where MC = MR, which in this example gives the price P_M. What we have added however in Figure 5.4 is the area labelled 'consumer surplus'. We partly examined this in the last unit, however this is the area where there are consumers that are prepared to pay more than the market price for the product.

One of the main options under monopoly is that the firm can attempt to 'capture' this consumer surplus for itself. The key would be to charge each consumer the exact price that they would be willing to pay for the product, in other words, divide the market up into different market segments.

MARKET SEGMENTATION

The price set by public transport operators, particularly those operating under monopoly market conditions, depends on "what the market will bear". Why this occurs is because the market does not consist of homogeneous customers. Different segments of the market respond differently to different changes in the attributes of a fare and a journey (e.g. comfort, flexibility, price, ease of changing reservation, level of advanced booking required). This is referred to as market segmentation.

Pricing policy under a monopoly thus relates to differing levels of elasticity. Both own price elasticity and cross price elasticity, examined in Unit 2, play an important role in pricing decisions, and can be used to an operator's best advantage. Travellers with an inelastic demand can be charged relatively higher fares, particularly in peak travel periods, and therefore contribute more to peak costs. In situations of elastic demand, relatively lower fares can encourage more travel.

There are key differences between national and local public transport regarding pricing policy, because market segmentation is generally not possible on local public transport, although a degree occurs with schemes such as concessionary fares. Nevertheless, market segmentation tends to be concerned only with train and airline travel. On local public transport, passengers tend to buy their tickets close to or at departure time, hence there is little opportunity to segment the market. Furthermore, most passengers on local public transport are sensitive above all else to waiting and journey times, both of which are independent of the fare.

PRICE DISCRIMINATION AND PRODUCT DIFFERENTIATION

The transport market will operate forms of both what is known as price discrimination and product differentiation to maximise revenue as part of yield management. Price discrimination is used to describe a pricing policy by which a firm discriminates between different groups of customers, who are charged a different price for identical units of supply. To put it more simply, we can say price discrimination occurs if two people have the same journey and package, but pay a different price. Product differentiation on the other hand occurs when different groups of customers are charged a different price for different units of supply. It usually involves different standards of service such as first and second class on trains or airlines, and can be a justification for price discrimination, particularly in terms of the image of the carrier.

Market conditions required for price discrimination

The following three market conditions are required for price discrimination to occur:

1. Monopoly control of the supply of the service. In order to practice price discrimination the firm must be the sole provider of the service. Not only must there be no current competition however, but also there must be no potential competition. Any price discrimination policy could be 'destroyed' by a new firm entering the market and charging the lower fare but making it available to high fare consumers. There must therefore be both monopoly control of supply and significant barriers to entry to ensure that this monopoly control will continue in the short and medium term.

2. It must be possible to divide the market into individual segments and thus separate different customers within a particular market. These segments must be clearly divisible and there must be no mechanism through which high yield premium charge consumers can 'down trade' to the lower charge segment. Such market dividers are known as inhibitors and prevent trading between different market segments from occurring. This tends to mean that price discrimination is more likely to be practised in service industries, such as transport, where the market can be divided relatively easily in terms of time. Indeed, divisions tend to be on the basis of time of day, day of the week or season of the year. Other inhibitors can be based upon geographical location, age and time of purchase (booking).

3. Differing elasticities of demand for each market segment. For each of the market sectors identified above, consumers must have different elasticities of demand. There is no point in dividing up the market into different segments if they all have identical elasticities of demand, as each segment would thus be charged the same price and hence there would be no price discrimination.

Example of consumer surplus

Let us return to the issue of consumer surplus. If the firm was to perfectly price discriminate, then it would be able to separate the market such that it would be able to charge each individual consumer the exact amount that they would be willing to pay for the good or service. In reality the only way of achieving this would be through a one-off auction where the items for sale are a limited number of highly valuable commodities, however such market conditions very rarely occur. In practice therefore, it is only possible to divide the market into a limited number of segments. To illustrate the principle, we consider the case of a train operator issuing train tickets from London to Bristol. The example concerns price discrimination against customers travelling on Fridays, and is completely hypothetical and simplified to show the underlying concepts. Figure 4b.5 illustrates the situation before price discrimination occurs.

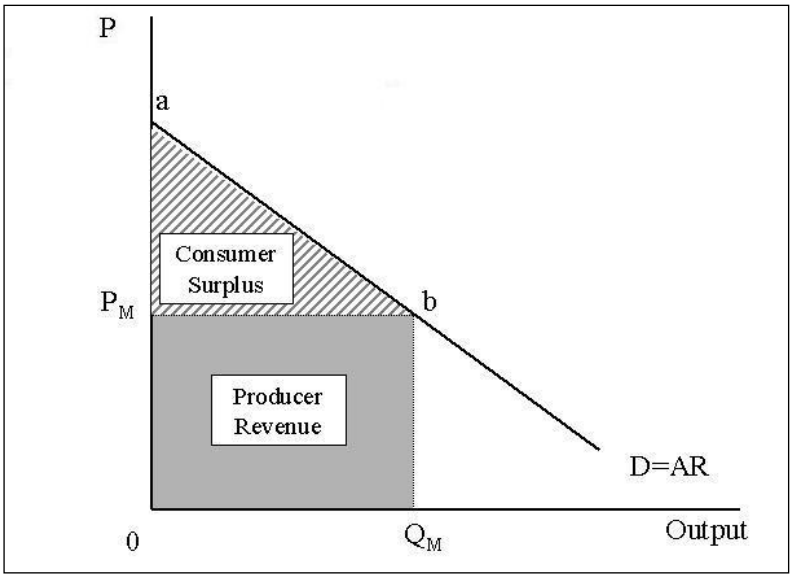


Figure 4b.5: Situation before price discrimination

This is our standard downward sloping demand curve, with the price charged by the monopolist being P_M, and the quantity supplied shown by the output level Q_M. We assume this is the profit maximising level of output before the practice of price discrimination. Revenue accruing to the producer is the grey shaded area outlined by 0, Q_M, b, P_M, which has an area given by the price of each unit times the quantity sold. The crosshatched triangular area shown above and outlined by P_M, b, a, is the area of consumer surplus.

If we now add some values to these figures and suppose that the price of a Normal saver fare for this rail journey is £20 (P_M) and the quantity sold on the journey is 9,000 tickets (Q_M), then the revenue before price discrimination is given by the simple formula of P_M x Q_M or £20 x 9,000 = £180,000

If the market is now to be divided into two market segments, then price discrimination can take place. The basis of this market segmentation would be on days of the week. If for example it is identified that 5,000 of these 9,000 passengers can only travel on a Friday and that they all value the service at a higher price than the current market price of £20, say £25, then this group can be charged the higher fare. This can occur because the firm has a monopoly control over supply and there is a clear inhibitor (travelling on a Friday). Note also that in this example we assume that others travelling on a Friday but that have alternative days on which they can travel will do so at the lower fare, hence these passengers are not lost to the operator through charging the higher Friday fare. If the Premium fare to be imposed therefore was £25, Figure 4b.6 shows the situation after price discrimination.

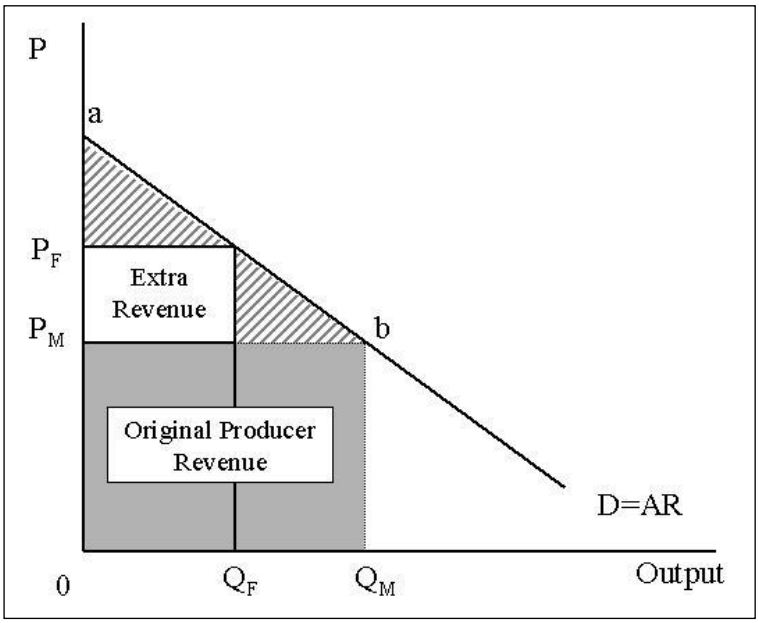


Figure 4b.6. After price discrimination

The market is now divided between those travelling on a Friday and those travelling on all other days of the week. If the Friday fare is set at £25, then the total revenue accruing to the operator is given by the not so simple formula of $(P_F \times Q_F) + [P_M \times (Q_M - Q_F)]$. In English, this translates as the price charged on a Friday (P_F) times the number of travellers on a Friday (Q_F) plus the price charged on all other days of the week (P_M) times the difference between the total market size (Q_M) and those travelling on a Friday (Q_F). Using our numbers, this would give a total revenue accruing to the operator of:

(£25 x 5000) + [20 x (9000 – 5000) = £125,000 + £80,000 = £205,000

This new total revenue of £205k represents a net increase of £25k. Price discrimination is possible in this example because the Friday (weekend away) market is less elastic than on other days. The market size is also greater and price discrimination may assist in reducing overcrowding. It may also prevent the increased costs of supplying additional train departures on the Friday for what are less price elastic consumers (as those consumers that can travel on other days now do so, hence using spare capacity on those days).

Price discrimination could also be introduced at the lower end of the market. This would be particularly pertinent in this example where the marginal cost of carrying these extra passengers would be extremely small, and hence any increase in revenue would virtually represent all profit. Additionally, new travellers could be enticed to make the journey by introducing lower fares, such as through the use of railcards. Train operators have to be

careful however that by introducing lower fares, those that currently use the train service do not downgrade.

Advantages and disadvantages of price discrimination

Advantages

- For the operator, the biggest advantage is that price discrimination increases the income that can be generated from a given service at very little cost (purely admin).
- If passenger demand is price elastic at the lower end of the price range, then a reduction at that level will generate new traffic from those who did not make a journey previously because they could only afford to travel at a lower fare. It therefore has aspects of equity attached to it, as it may bring such transport services more into the range of those on lower incomes, thus giving greater choice to a larger proportion of society.
- Further social advantages can be obtained by providing cheaper travel for lower income groups e.g. families, elderly people. In particular this may increase mobility and accessibility, both of which are key aspects of the social inclusion agenda.
- Resource cost savings for the economy as a whole may be achieved as trains are now run under a franchise contract in which they are required to provide a service irrespective of demand. Any spare capacity can thus be used to price based on the marginal cost and hence produce a modal shift from car to train and this will reduce congestion (and its associated costs), produce fuel savings and give wider environmental benefits.
- An operator may be able to make a profit out of a transport service through price discrimination, whilst if forced to charge a single price the service would make a loss and hence not be provided.

Disadvantages

- For the operator, a customer may move from a high yield, first-class, executive passenger to a low yield, second-class, saver fare passenger. To counter this, an operator will place restrictions, such as time restrictions, and this underlines the importance of strong inhibitors.
- If not properly managed, an increase in demand in the discount price segment can create a peak in demand that results in overcrowding or requires the provision of an additional vehicle to meet the demand. This would therefore result in increased costs.

- Passenger dissatisfaction may result from price discrimination where one passenger pays the standard fare and another travels at a discount, but both experience exactly the same standard of service.
- Passengers whose demand is inelastic may be disadvantaged by price discrimination because they have to pay premium fares to arrive at work, or at business destinations on time.
- A reduction in consumer surplus. Most of these issues have been concerned with the operator, but one of the biggest disadvantages and main arguments against price discrimination is that it reduces the size of the consumer surplus. Consumers therefore have to pay a higher price hence have less income left over for other goods and services. In the case of transport, this problem is particularly acute, as most individuals have to travel to earn a living, and for those on lower incomes, the percentage of income spent on travel is far higher. Price discrimination therefore, can be regressive, as it is taking income from low-income groups (those on low pay) and giving that in the form of the profits of transport operators to higher income groups (shareholders). It is therefore economically inefficient.

Before we leave the issue of pricing in theory, a summary of the three types of market structures is provided in the appendix to this unit. You are referred to this table for a succinct overview of the major differences between these three market types.

Pricing in Practice

We end this unit by having a very brief look at pricing in practice under the sub-headings of ticketing and fare structures.

TICKETING

Ticketing is mainly covered in the Public Transport module, hence we only very briefly touch upon it here. Tickets that are easy to buy, offer value for money and are flexible encourage more people to use public transport. Integrated tickets are also helpful to travellers because they allow them to change operator or even mode within a specified area. The London Oyster card is one of the best examples in Britain of a successful, integrated area-wide ticket scheme. The card provides unlimited pre-paid travel within specified zones on bus, rail, underground and Docklands Light Railway services throughout the capital.

A similar ticket scheme operates around the Glasgow conurbation. The Strathclyde PTE (the body responsible for public transport information and integration around Glasgow) offers a SPT Zonecard. All the major and most of the minor bus companies operating in and around Glasgow have signed up to the scheme, together with Scotrail. The ticket is also valid on the Glasgow Underground. It is important to highlight however that such (multi-modal) tickets general are considerably more expensive than single operator cards. In Glasgow for example, the 4 week Zonecard is roughly £10 more expensive than a Firstcard (which can only be used on Firstbuses) covering a similar area size. This underlines that the structure of bus fares, indeed all public transport fares, outside London can be very complex. Without an integrated system, there is unnecessary delay at bus stops whilst fares are being collected. Passengers who have to change buses in the course of their journey usually have to pay twice, spending more than if they were making the journey on one bus.

FARE STRUCTURES

The simplest fare scale is a flat fare, whereby each passenger pays the same fare each time a vehicle is boarded, and can travel for any distance. Flat fares are rare in the UK (one example is the Glasgow Underground) but common elsewhere in the world. On mainland Europe there is another key difference, that the metric used is time rather than distance. For example, a single fare may well buy travel for up to 60 or 90 minutes from the time the ticket is purchased. The operator has a difficult balancing act with a flat fare scheme. If they price it too high then short-distance passengers might not travel, but if they price it too low then long-distance passengers have a great bargain.

The convention in the UK is for a graduated fare, which is illustrated in Figure 4b.7. A feature of graduated fare scales is that the fare is not in linear proportion to distance, but has an initial step (boarding charge), followed by a given rate of increase thereafter (a taper). Therefore, the fare per kilometre reduces as the journey length increases. To some extent this is a reflection of costs, as each passenger boarding incurs costs of ticket issue and delay to the vehicle. However, it has often developed in a rather accidental fashion, as successive fare increases have caused a higher percentage to be applied to the lower fares.

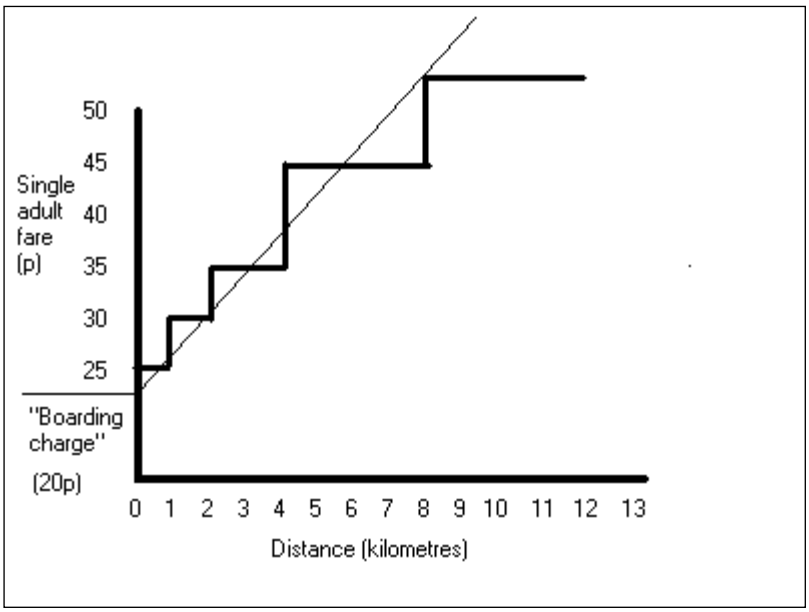


Figure 4b.7. An example of a graduated fare scale.

Generally stops are grouped in fare stages, typically 1 kilometre long. Many local bus operators charge their fares based on the number of stages. There may be an element in the fare incorporating time of day. For example, on Lothian Buses you pay £2.00 rather than £2.50 for a Day Saver ticket if bought after 9:30 on weekdays (off-peak). As an example of local bus fares, Lothian Buses had the following fare structure introduced on 12th March 2000:

Adult Single Fares	1-2 stages	50p
	3-8 stages	80p
	9-13 stages	90p
	14+ stages	£1.00
Child Single Fare	Any distance	50p

OAP Single Fare	Any distance	30p

Interesting this fare structure was subsequently revised, hence as at 4th January 2005 Lothian Buses fare structure was as follows:

Adult Single Fares	1-8 stages	80p
	9+ stages	£1.00
Child Single Fare	Any distance	60p
OAP Single Fare	Any distance	40p (before 9:30, free after 9:30)

This was revised again in 2007 (I think), to a single flat fare of £1.00 (now £1.20 at the time of writing).

There is a general trend to simplifying fare structures and moving towards flat fares. From an economist's point of view however, this will not lead to market efficiency, i.e. consumers making the right choices. This is for the simple reason that if the fare is not connected to the cost of providing the journey, as is the case with flat fares, then the link between the benefits gained from consuming a good or service (as measured by the price paid) and the cost of providing it is lost. Market efficiency occurs where the price equals the marginal cost. It will not be explained here why this is the case, as at this point in your study of transport economics you should be able to work that out for yourself (big hint – the P = MC requirement assumes the market is in perfect competition). It does suggest however that transport markets are moving away from a position from which they could achieve market efficiency. The link between pricing transport services in practice and pricing them in theory however is a much under-researched area, and one that is worthy of far more attention. For example, the underlying assumption of consumers is that they are rational and that they seek to maximise utility. Certainly the first of these is sometimes questionable, however with the second this may be partly gained by simplifying the fare structure facing the consumer, i.e. is a psychological factor. An example of this is given below.

EXAMPLES OF TRAIN FARES

The following exert, from the UK Integrated Transport White Paper[15], shows the then situation with regard to train fares:

[15] Department of the Environment, Transport and the Regions (1998), *A New Deal for Transport: Better for Everyone, The Government's White Paper on the Future of Transport*, (*http://www.detr.gov.uk/itwp*).

One of the most obvious failures of rail privatisation has been the perceived lack of a clear, understandable national fare structure. Some key fares are regulated by the Franchising Director and from 1999 until 2003 fare increases will be restricted to Retail Price Index minus 1% - a fall in real terms[16]. But many popular fares such as APEX, cheap day singles and returns are set entirely at the discretion of the individual train operator. Although train operators have introduced some new and innovative fares, this has led to a multiplicity of different and frequently changing fares for similar services with, in some cases, complex and varied conditions, for example in relation to advance booking.

The controls, and the absence of controls, are a consequence of legally binding franchise agreements inherited from the previous Government. There is little practical scope for altering them in the short term. But when opportunities arise for negotiating franchises, the new Strategic Rail Authority, guided by Ministers, will ensure that arrangements are made so that train operators structure and market their fares to offer value for money for their customers. The SRA will also ensure that fares reflect the fact that the railway is a national network which needs to be marketed accordingly and in a way that encourages people to switch from car to train.

An internet search of fares from the GNER (Great North Eastern Railway) website (www.gner.co.uk) back in May 2000 revealed there to be 23 different tickets available, not including the difference between single and return tickets. The tickets available were:

First class (6): First class open single/return, Executive day return, Executive day single, Executive special package, Weekend first, Leisure first return.

Standard accommodation (10): Standard open single/return, Standard day single/return, Saver single/return, Supersaver single/return, Super advanced single/return, Bargain single/return, Apex single/return, Daypex, Fab 4, Season tickets.

Promotions (3): Fab 4 Leeds/ London, Fab 4 Scotland, First Stop York.

Railcards (4): Young persons railcard, Discount railcard for people with disabilities, Family railcard, Senior railcard.

Therefore, a large range of train tickets were offered by GNER, and the price differences could be explained as follows:

- First class fares offer the customer extras such as food and car parking;

[16] As we will see later, you should note that regulated rail fares are now regulated on the basis of RPI+1%, i.e. a rise in real terms.

- Special up-grade tickets to First Class are available in the off-peak times (i.e. the weekend, when business travellers tend not to travel);
- Fares are cheaper (i.e. not Open) if it is specified when the customer makes the return journey;
- Fares are cheaper (Saver) if the customer doesn't travel in the daily peak to London;
- Fares are even cheaper (SuperSaver) if the customer doesn't travel in the weekly peak (Friday);
- Fares are cheaper when purchased in advance (APEX);
- The promotions (Fab4) give discounts to groups of customers travelling to low demand destinations during low demand periods; and
- The railcards give the chance for lower income groups (i.e. elderly) to travel by train who wouldn't otherwise use the train.

A search of the Trainline ticket service for services/fares in November 2009 however, revealed a very different picture – 8 tickets were available for an open return and 6 for a fixed return. Of the latter category, these consisted of the cheapest single category, cheapest first class single, super off peak return, off peak return, standard return and cheapest first class return. This is probably all that is required to discriminate in the market and also underlines the move towards simplified fare structures, although in this case note that this is as a result of less price discrimination (although just as effective) rather than a move towards flat fares.

Unit 4b Self-assessment exercises

Exercise 4b.1

Briefly summarise the major similarities and differences between an oligopoly and monopoly market structures. Then illustrate, without reference to the notes, the profit maximising position for an oligopolistic market.

Exercise 4b.2

Refer back to self-assessment exercise 4.2. On your 'degree of competition' scale, where would you place oligopoly? Of the nine industries/markets listed in that exercise, which three do you believe most closely resemble an oligopolistic market structure? Why have you chosen these three?

Unit 4b Summary

In this unit we have looked at the key economic concepts of pricing under different market structures, yield management and some examples of pricing in practice. These concepts are valuable to both public transport operators and those that study these industries, whether that be students, academics or regulatory authorities. We introduced the idea of the oligopoly market structure and looked at the issue of price leadership. We then defined price discrimination and product differentiation under monopoly market conditions and a working example was provided on price discrimination on the trains. Consumer surplus was also introduced as a concept and illustrated using graphs.

For the remainder of the Unit, the focus was in ticketing. We looked at the two main fare structures, flat and graduated fares, whilst examining the range and type of tickets available on public transport.

Appendix to Unit 4b – Summary of Main Market Structures

Table 4b.A1. The different market structures.

Type of market	Number of firms	Freedom of entry	Nature of product	Examples	Implication for demand curve for firm
Perfect competition	Very many	Unrestricted	Homogeneous (undifferentiated)	Agricultural products, truck load road haulage (these approximate to perfect competition)	Horizontal. The firm is a price taker
Oligopoly	Few	Restricted	1. Undifferentiated or 2. Differentiated	1. Buses, Airlines, Rail Freight 2. Cars, Low cost airlines	Downward sloping, relatively inelastic but depends on reactions of rivals to a price change
Monopoly	One	Restricted or completely blocked	Unique	Local train operating company, bus services in London	Downward sloping, more inelastic than oligopoly. The firm has considerable control over price

Source: Adapted from Sloman (2009), Economics, 7th Edition

UNIT FIVE[17]
THE REGULATION OF TRANSPORT SERVICES

INTRODUCTION

In many respects this unit is concerned with economic 'control' and specifically the control by relevant transport authorities on the levels and behaviour of transport users and operators under their authority. In this unit we will concern ourselves with economic regulation, sometimes referred to as quantitative regulation, however you should also be aware of qualitative regulation, which basically lays down the laws of operation of a given transport network. We begin by examining the different forms of regulation before very briefly looking at the ownership structures of transport services (as public ownership is an alternative to economic regulation).

RECOMMENDED READING

The recommend reading for this unit is:

Cowie, J. (2010), The Economics of Transport, Chapter 10, Transport Regulation and Ownership, which is still worth reading despite the footnote!

Mallard, G. and S. Glaister (2008), Transport Economics, Chapter 12, Privatisation and Deregulation (please note however that this chapter is very one dimensional)

LEARNING OUTCOMES

Once you have worked your way through this unit you should be able to:

- Outline the rationale behind the economic regulation of transport services
- Understand the forms that this regulation can take
- Give an overview of the forms of public ownership that exist in transport markets.

[17] Please note that this is not Napier copyright but rather owned by Taylor and Francis, as large segments are plagiarised from Cowie (2010) (and yes, believe it or not but you can plagiarise your own work!). However I guess that's ok under the CLA (the copyright agreement between publishers and UK universities).

FORMS OF TRANSPORT ECONOMIC REGULATION

Quantitative regulation, more commonly known as economic regulation, is where the regulatory bodies intervene in the market in order to place economic controls on the operation of the market. This is either in terms of restrictions with regard to the price or restrictions or minimum specifications with regard to the supply. Dealing firstly with price regulation, this comes in a number of different forms:

SPECIFY THE PRICE TO BE CHARGED.

In theory, this is the simplest form of price regulation and is illustrated in Figure 5.1.

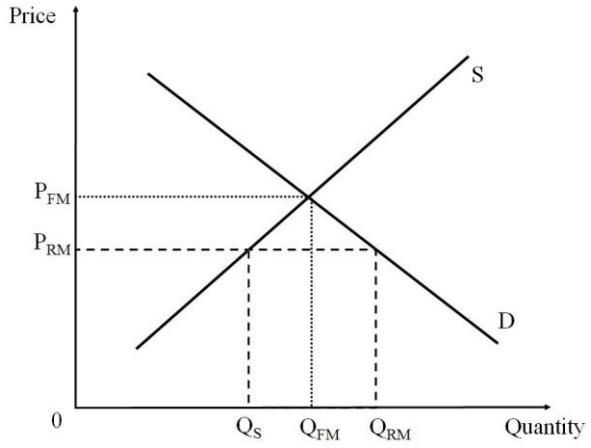


Figure 5.1: Simple price regulation

In Figure 5.1, the free market price is given where supply equals demand, hence P_{FM}. The transport authority however sets a maximum price below the equilibrium price at P_{RM}. Note it only makes sense to set a regulated maximum price below the equilibrium price, otherwise it would have absolutely no impact on the market. As can also be seen from Figure 5.1 however, a price at the regulated fare will create excess demand i.e. there will be more wanting to use the service than is being supplied to the market. This is shown by the difference between Q_{RM} and Q_S. The authority would therefore have to address this problem. What it would have to ensure is that the supply curve moved to the right. This can be done in several ways, with the 'simple' solution being to pay the operator a subsidy. Alternatively the whole rationale for imposing a maximum fare may be to motivate the operator to lower its cost. Hence the supply curve would shift to the right thereby eradicating the excess demand. In practice this has normally been done over time by regulating price increases, as outlined below.

SPECIFY THE MAXIMUM INCREASE IN PRICE ALLOWED

Rather than state a specific price, the authority limits the extent to which the operator can increases it prices over time. In the UK this has normally been done through an 'RPI-X%' formula, where RPI relates to the prevailing rate of inflation and X the value to which price increases are restricted to. Hence in theory prices should rise at lower than the rate of inflation, thereby falling in real terms. Hence if the rate of inflation is 3 per cent, and X is set at 2 per cent, then the maximum fares can rise by is 1 per cent. In real terms therefore, this would be a 2 per cent cut in the fare. As stated above, this would primarily be used to motivate efficiency improvements, as it is only through reducing costs, and not increasing prices, that operators can maintain or increase profitability.

Nevertheless, where price increases are expected to rise by well in excess of the rate of inflation, these can be limited through an 'RPI+X%' formula. Thus fares on most of the rail franchises in Britain are currently regulated on the basis of RPI+1% with some of the franchises allowed to raise fares by RPI+3%.

REGULATE THE (FINAL) PRICE BY TAXING THE GOOD OR SERVICE

Varying tax levels can be used to regulate the price in the market, and these can be specified either as a percentage or set at a specific value. In most case such taxes are used as general tax raising measures in which the government acquires public finance to spend on the provision of public services. In the UK for example, all goods except exempt items are charged Value Added Tax (VAT) at a rate of 17.5 per cent. Hence this adds 17.5 per cent to the price of the good with this additional tax revenue passed on to the government. Additional or specific taxes however may be imposed to regulate the price in the market, thus fuel duty has the effect of significantly increasing the price of fuel to the consumer. If the government wished to limit car usage, it could do so by increasing fuel duty, thus increasing the price. This would be known as a 'Pigouvian' tax, which is one that is imposed in order to correct for a negative externality.

SPECIFY THE RATE OF RETURN (PROFIT) TO BE GAINED.

Prices charged by transport operators can be regulated based upon the level of profit to be gained. Hence a 'reasonable' rate of return may be set and then prices regulated accordingly to achieve that rate of return. This will normally take place where the level of demand can be estimated to a very high level of accuracy, and hence the only real variation in total revenue will be as a result of the price charged. Network Rail for example is regulated by the Office of Rail Regulation with regard to the level of track access charges it imposes on the train and freight operating companies on this basis. The charge is based upon a rate of return on the assets employed. Given that the level of train movements on the network are known a year in advance, then the level of revenue gained can be accurately predicted by the price. This however is a considerable oversimplification of the problem, as obviously operating costs as well as efficiency gains

also need to be estimated in advance. If efficiency gains are not taken into account, then the level of the access charge, or fare in more general terms, would be set at a rate that would provide a higher level of profits than had been deemed to be a 'fair' return.

THROUGH INTRODUCING YARDSTICK COMPETITION

Yardstick competition exists where direct competition in the market is not feasible but is introduced indirectly through regulation, and is normally used to control price levels. This is achieved by linking the performance of different firms in different markets to each other. If it was conceived today therefore, 'yardstick' competition would probably be termed 'benchmark' competition, as effectively the performance of each firm in the industry is benchmarked against one another. As an example, if two firms A and B faced similar cost and market conditions, then under yardstick competition the price that A could charge would be dependent upon the level of costs in firm B. Thus if B was to lower its costs, the result would be that under the regulatory system A would be forced into charging a lower price in its market and vice-versa. Such regulatory measures are said to be appropriate where the potential for cost savings are unknown and difficult to estimate in either market, and where cost and demand conditions in each market are very similar.

These are the five main mechanisms through which the price charged in the market can be controlled, however price regulation is only one form of economic regulation. The other main form is where the level of available capacity is controlled through regulatory measures, normally in the form of specification of minimum frequencies and/or through the control of market entry

SPECIFY A MINIMUM FREQUENCY:

In the simple case the authority specifies the minimum level of frequency to be provided. This will normally be in the form of actual frequencies and operating hours, however it can take other forms such as total vehicle kilometres supplied. This is shown below in Figure 5.2.

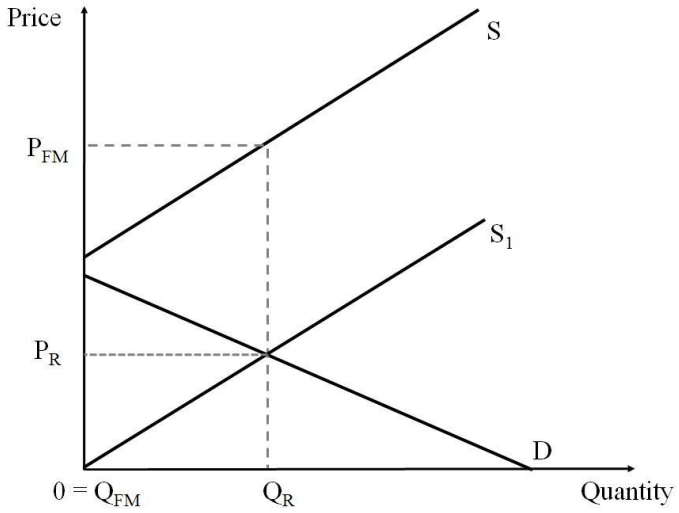


Figure 5.2: Simple quantity regulation

In Figure 5.2 the free market position is shown by supply curve S and demand curve D. In this case no market would exist for this transport service as the highest price that consumers are willing to pay, as shown by the demand curve, is below the lowest price that suppliers would be willing to provide a service, as shown by the supply curve. There is no market equilibrium and thus no transport service is provided. In this case, the transport authority specifies a minimum service frequency in order to create a regulated market where none existed before. This is set at level Q_R. Note however that such action taken on its own would be entirely pointless; even if an operator did, for whatever unwise reason, enter the market it would quickly go out of business as the level of demand would be insufficient to support the level of supply. As with simple price regulation therefore, such regulatory action cannot be taken on its own. The regulatory authority would need some belief that supply will increase and result in a shift of the supply curve from S to S_1. Thus it may again pay a subsidy or alternatively 'package' the non-existent service with other more profitable routes and put these out to tender to the highest bidder.

LIMIT MARKET ENTRY

We have already seen in unit 4 that legal control of market entry is a barrier to entry, but it is one that would usually be imposed in transport markets as a form of capacity regulation. This is where the regulatory authority will set clear limits on those operating in the market and thereby directly restrict supply. Taxi services in most locations are a good example of this form of regulation, where the regulating authority sets a limit on the number of licences issued. Such measures were originally imposed to avoid street congestion and to limit the competition to public transport services.

THE RATIONALE FOR THE REGULATION OF TRANSPORT SERVICES

In most cases regulations, both qualitative and quantitative, are imposed due to some form of market failure, with the main reasons outlined below.

To overcome the market failure of imperfect/dissymmetry of information

The first rationale for the regulation of transport services covers most forms of qualitative regulation, where these are imposed in order to regulate user behaviour and to impose minimum standards on the operation of the transport system. This is directly to overcome the market failure of imperfect information, or more generally because a dissymmetry of information exists where some know more than others. In many ways this is akin to the view that every other driver on the road is an idiot apart from me, and in many ways, there is actually some truth in that statement. In general, users of a transport system need to have a level of confidence in others' movements in order to feel more secure when undertaking their own. Hence when walking on the pavement, pedestrians need to feel reasonably assured they will not be run over by a high speed vehicle. Similarly, drivers need to feel reasonably assured that pedestrians will not aimlessly walk out in front of the vehicle hence causing them to swerve dangerously or worse, run them over! They also need to have confidence in the fact that other drivers will not attempt to use the same piece of road space they are using at that point in time hence causing a collision. Through regulating behaviour, e.g. you can only walk on pavements, you must obey road signs, you must adhere to maximum speed limits etc, a minimum overall level of behaviour is assured when moving around and total chaos avoided.

Not only individual behaviour however needs to be regulated, but also the condition of vehicles used in the transportation of people and freight. This relates to all forms of transport, private, public and freight haulage, and through this a minimum standard of the equipment used is ensured. None of us when boarding a bus, or an airline, or a train, have sufficient knowledge to be able to assess the skills of the driver/pilot or the condition of the vehicle. Even if we did, the practicalities of doing so would make the whole system completely unworkable. Such knowledge however is not necessary, as that is one of the main functions of the regulatory authorities. They have the skills and the knowledge to assess these issues on behalf of all users of the transport service. In simple terms, qualitative regulation overcomes the market failure of imperfect information and makes the whole system work.

The market can no longer regulate itself

If the market can no longer regulate itself, then external regulation is required to ensure that economic efficiency is achieved. Why this occurs is because most transport industries tend towards anti-competitive market structures, namely monopoly and oligopoly, hence regulation is required to attempt to minimise the disadvantages associated with such market structures.

To correct for externalities

Even where the market can regulate itself through strong competitive pressures, it may still not produce the 'right' answer in terms of modal splits that maximise economic welfare. This is due to the high level of externalities present in transport markets, hence all decisions are based on private costs and benefits and do not take into consideration the wider implications of these decisions. As a result the 'wrong' society maximising decision is reached. External intervention in the form of regulation, which although a further breech of the conditions of perfect competition, is required to rectify this situation on the basis that two 'wrongs' will make one 'right'. The extent to which this has been successful or not in actual transport markets is considered later in the chapter.

To ensure the quality of the service provided

Ensuring the quality of service provided in the market relates directly to qualitative controls on the standards of service to be provided. Whilst relating to 'qualitative' controls however, most of these are imposed through economic and not qualitative regulatory measures. Thus for example in Britain, Statutory Bus Quality Partnerships are agreements between the local authority and the bus operators, where the local authority can limit entry to the market to only those bus companies that meet the vehicle standards specified in the agreement. Other examples exist where the quality of service is regulated even more directly. British passenger rail franchises agreements for example specify the minimum level of service to be provided and the investments in new rolling stock to be made by the operator over the course of the franchise.

Regulation may also be used to ensure the quality of service in the longer term. For example market entry may be restricted in order to provide market protection and hence maintain the profitability of those already in the market. It is out of profits that investments are funded; hence restricting entry gives current operators the business confidence to make future investments in new vehicle stock and thus maintain or improve quality levels. A similar reason may be to protect the livelihood of those in the industry. The most well known example is the EU's Common Agricultural Policy, which was originally conceived to lend support to farmers within the industry through regulating the prices of agricultural produce. Regulated prices were set above the market price in order to ensure future livelihoods and thus the future of the industry.

To provide a transport service where none existed before

Rather than leave entirely to the free market, transport authorities may decide to intervene and package routes in order to ensure that all necessary services are provided. It may therefore restrict entry on certain (profitable) routes in exchange for the protected operator providing services on unprofitable routes. This is known as cross subsidisation, where the revenue earned from profitable services is used to fund the losses incurred on unprofitable routes. Whilst such measures prevent the cherry picking of 'good' routes and can provide a far more overall balance in the provision of transport services, there are generally far better measures that can be used to ensure that all necessary services are provided. This is a topic that is developed further in the next chapter.

To improve efficiency within the industry

As outlined above, the regulatory framework can be used in an attempt to bring about efficiency improvements within the industry, normally through restraining price increases. This was a measure that was extensively applied in the British privatisation programme of the 1980s, where charges relating to telecommunications, electricity, gas, water and so on were carefully monitored by an industry specific regulator. After privatisation, rail fares in Britain were also regulated on an RPI – X% basis, where X first equalled zero and then subsequently one. These measures are used to not only instigate efficiency improvements, but also to ensure that productivity improvements result in lower consumer prices rather than higher shareholder dividends. As noted above however, the British rail franchises are now regulated on an RPI+X% formula, which hardly provides an incentive to improve efficiency but has been introduced for other reasons mainly surrounding investments in rolling stock.

THE DRAWBACKS OF ECONOMIC REGULATION

We have looked above at the rationale behind the regulation of transport activities, however it does not always result in the desired outcomes and has a number of drawbacks.

Limits free enterprise.

Often cited as the biggest drawback of economic regulation, the issue of the limitation of free enterprise was often put forward as a reason for the removal of economic regulatory measures during the enterprise culture of the 1980s. Acting out of self interest, whether that be at a company or at an individual level, is said to be a far stronger motivator to 'do the right thing' than a regulator acting in the public interest. This is related to the free market ideology of consumer sovereignty, where those that profit most are those that are able to give the consumer what they want. In a transport context, the argument would be that not only would an entrepreneur be far better positioned to identify users' needs and provide services users want, but also would be far more motivated to do so by the profit motive, i.e. direct benefit to themselves. This contrasts with a regulatory authority that would plan such networks at a distance and be motivated by the public interest, i.e. simply doing the job they are employed to do. Closely related to limiting free enterprise, regulated markets are said to also limit innovation, as it dampens the free enterprise spirit. This is because there are clear limitations imposed upon the market and consequently less room for the operator to use entrepreneurial flair and innovative solutions in the provision of transport services.

Inefficient, second best solution

As will be developed further in the next chapter on subsidy, issues of subsidy and regulation are what are termed 'second best solutions'. The best solution is that the market itself regulates the performance of operators. This would be an internal/automatic type of regulation. As we saw in Unit 4 with perfect competition, any operator that cannot offer services at the lowest price and produce these at the lowest possible cost will be

driven out of business, hence the market regulates itself. The problem with an external regulatory body is that this leads to added costs in the operation of the market, as the administrative burden of the regulatory mechanism adds to the overall cost of providing the service. The annual costs of running the Office of Rail Regulation (ORR) in the UK for example came to a total of £30.5m in financial year 2007/8 (ORR, 2008). Whilst these are not all additional costs, as for example if the ORR did not exist then a number of its duties would need to be amalgamated into other government bodies, a large part of it is, particularly those concerning economic regulation. Thus transport services are provided at a higher cost due to increased administration costs.

Also under this heading is the time issue, and in particular the time gap between changes in market conditions and changes in the provision of services. It has often been argued that due to the added layer of regulatory bureaucracy, a regulated market does not act as quickly as the free market in responding to changing conditions in demand and supply. A classic case of this would be the bus industry in Britain, where during the period of the bus wars a large number of mergers and acquisitions were referred to the relevant regulatory body, in this case the Monopolies and Mergers Commission[18]. By that time however, the act of merger/acquisitions had already occurred, hence many of these referrals ended in purely nominal measures and the upholding of the merger. Cowie (2002) for example highlights that of 33 cases referred to the Monopolies and Mergers Commission between 1986 and 1998, only 3 resulted in dissolution of the merged company.

Asymmetry of Information

In order to regulate efficiently, the regulator needs a high level of information in order to plan and control operations. Whilst sounding obvious, the major drawback with this is that the regulator will undoubtedly be the authority and not the operator, and hence the operator, unsurprisingly, knows more about their own business than the regulator does. There is therefore a need for a flow of high quality information between the two. It may however be in the interests of the operator to withhold important information from the regulator if they believe this may be used against them. There is thus an imbalance of information, with the operator often holding the key information that the regulator needs to regulate effectively.

The issue of regulatory capture

The theory of regulatory capture was originally put forward by George Stigler (1971) and after the negative effect on entrepreneurial flair is often cited as the second biggest drawback of economic regulation. The issue of regulatory capture is where the regulator, who is supposed to oversee and control the industry on the grounds of the public interest, is in effect not as 'tough' on the industry as they perhaps should be in the public interest. The consequences are that the regulator better serves the interests of the industry than the interests of the consumer. Whilst the term 'capture' suggests that the regulator is

[18] The Monopoly and Mergers Commission was the predecessor of the Competition Commission, which took over its functions in 1999.

enticed in some manner, 'capture' also refers to capture with regard to the prestige that goes with being an industry regulator. In simple terms, the status and overall prestige of the regulator rises with the importance of the industry. What happens therefore is that instead of regulating the industry, the regulator becomes the protector of the industry. A more direct form of regulatory capture is where the regulator becomes dominated by the vested interests in the industry, where political pressure may be applied from above forcing the regulator to take a 'soft' stance in regulation of the industry.

Cumbersome regulatory procedures make avoidance of regulatory measures possible

This is the most basic drawback of regulation where it simply fails to regulate the actions or behaviour it is designed to regulate through avoidance. This is because regulatory measures are by their very nature cumbersome processes that can be avoided or ignored by those they are designed to effect. As a result, the expected outcomes are not achieved. It can be very difficult and costly to regulate an industry, and such cumbersome processes can make it very difficult to get it 'right'.

Note that whilst these major drawbacks of regulation have been listed separately, they all in some form relate to the issue of it being a second best solution – in simple terms, the only problem with regulation is that it is needed at all. If there are thus problems with regulating the market for transport services, the alternative is to take it into public ownership.

THE ISSUE OF OWNERSHIP IN TRANSPORT MARKETS

Transport markets with one or two exceptions cannot be left entirely to market forces to resolve economic transport issues. In most cases therefore, they need some form of external intervention in order to correct for market failures. One way that this can be achieved, as seen above, is through regulation of the market. The alternative is far more direct and is where the state effectively takes control of the market by bringing it into state ownership. The market therefore does not have to follow market principles, i.e. be subject to the forces of supply and demand, nevertheless the economics of the whole operation still need to add up.

Reasons for the Public Ownership of Transport Assets

The issue therefore of ownership within transport markets in most cases comes down to the simple choice of whether services should be provided by the public sector or a regulated private sector. The 'old' view was that transport services, both passenger and freight, were vital services to the national and local economies that they served. Rather than questions surrounding issues such as the standard of service to be provided or the fare to be charged being left to the private sector to decide, these should be resolved by the state on the basis of the public interest. The 'best' way to achieve this was through direct ownership and control of the assets required to produce such services. This view is best exemplified with what eventually developed into the 'Morrisonian model' of public

ownership, named after one of the leading Labour politicians of the time, Herbert Morrison. In the early 1930s, London's public transport services comprised of buses, trams and the underground, and all operated as individual transport modes by 89 mainly private sector operators. Morrison's view was that these individual services could only be fully exploited (in terms of public benefit, not profitable reward) as part of a single unified system that was controlled and operated by a single transport authority. The authority should operate 'at arm's length' to the local authority with the remit to provide economic and efficient transport services. This was the origins of London Transport, which was enacted by the London Passenger Transport Act (1933) that established the London Passenger Transport Board to bring all passenger modes in London into public ownership and under the Board's direct control.

However, the public interest argument is not the only reason why assets should be taken into the public sector, or nationalised, the main other ones being:

Eradicate wasteful competition

Wasteful competition has already been outlined in Chapter 7, and is where two or more services exist where one would be sufficient. One way that this can be eradicated is to bring competing services under the control of a single operator. The problem however with such a re-organisation is that the eradication of competition would leave a monopoly provider with all the associated problems of such a market structure. One way these drawbacks can be overcome is through direct control of the operator via public ownership. The operator could then be managed on the basis of the public rather than the profit interest. If managed correctly, this should result in economic efficiency. Alternatively, wasteful competition could be removed through regulation, thereby replacing the free market with a planned or regulated market and ensuring that capacity matched demand.

Military significance

The industry is seen to have important military applications that mean it cannot be simply left in private hands. In some ways the rail industry is an example of this, where during WW2 the 'big four' rail companies were jointly managed in order to better assist the British 'war effort' and to co-ordinate troop and equipment movements.

Public goods

If left to the market certain types of beneficial goods, known as public goods (see Chapter 1 for a description) would not be provided as no single firm could make a profit out of doing so. One way to ensure that such goods are provided is through state ownership.

Essential to the economy

This is an extension of the public interest argument. Certain industries were viewed as underpinning the whole economy and hence could not be left to market forces because if the particularly industry suffered the whole economy would suffer. Hence for example the coal industry was nationalised in 1947 and steel in 1965 as at that time these were important raw materials to the whole British economy. Many transport industries came

under this same argument, hence the railways nationalised in 1947, road haulage and the major ports in 1948 and most of the bus industry taken into public ownership in 1968.

A large employer

Very much a rationale of the times of mass production, where firms were major employers and thus their potential collapse could not be contemplated in unemployment terms. As an alternative, they were nationalised. Hence British Leyland, a major British car manufacturer and mass employer at the time, was nationalised in 1975 following its bankruptcy. This was and remains a particularly acute threat in certain areas or localities, where certain firms may be the only major employer in the area, and hence its collapse would have implications far beyond the direct loss of employment. The aerospace industry in South West England for example is a major employer, and hence the collapse of the industry would result in considerable economic hardship throughout the whole region.

Key industry

A key industry is different from an essential industry to the economy as it is one that is seen to be of vital importance to the country. Hence for example Rolls Royce was nationalised in 1971, the only firm to be taken into the public sector by a Conservative government, for reasons that primarily related to it being a key industry. This is because firstly retention of the skills and knowledge employed in the company were seen to be important to the country and secondly, it was and remains one of a small number of 'prestigious' companies that stand for excellence in British engineering. Note also that the Rolls Royce example falls under the military rationale, as most Royal Air Force and Royal Navy aircraft are powered by Rolls Royce engines. Retention of such engineering skills and knowledge was therefore also of military significance.

High project development costs

Any major project requires considerable financial outgoings before any revenue is forthcoming. In many instances these can be of such a size that the whole company is put at risk. In the case of Rolls Royce for example, the company ran into major cash flow problems over the development of the civil RB211 aero engine that ultimately forced it into liquidation. Because of such high business risks, projects of this nature would not be undertaken by the private sector regardless of any wider benefits that may follow completion. Consequently under 'normal' circumstances, only a publicly owned company could take on such a project.

In the 21st Century many of these arguments are outdated and reflect different times. In several cases, other alternatives to public ownership can now be used to overcome many of the issues. In other instances however, reasons for public ownership still apply. For example, the net outcome of the reform of ferry services to the islands off the west coast of Scotland, which was required under EU competition laws, resulted in the services being provided by a combination of two heavily subsidised public sector companies rather

than previously where they were only provided by one heavily subsidised public sector company. All private sector bidders withdrew in the course of the bidding process as they were unwilling to bid under the terms of the tender. As the ferries are essential to the economy and social foundation of these islands, if no private sector company was willing to take the risk in providing services they could only be operated by a public sector concern. Whilst this is not a typical example, it does nevertheless lead into the more general issue of public ownership reform.

REASONS FOR REFORM

If transport services should or can only be provided on the basis of the public interest, which can be 'best' guaranteed under the direct control of public ownership, then why reform? Some of the reasons for reform are outlined below:

Increasing discontent with the model of public ownership

Over the years there has been increasing discontentment with the model of public ownership for a number of reasons. The first is the basic efficiency argument, where such organisations are perceived to suffer from general management slackness and thus are not as efficient as private enterprise. This is the classic x-inefficiency argument. The second relates to the constraints that operating in the public sector imposes on such concerns, which may limit options and overall performance. This particularly surrounds financial constraints, where because all borrowings count as part of the National Debt, these are closely scrutinised and controlled. For companies and organisations working in the public sector this limits planning horizons and investment plans, and consequently leads to much shorter planning frameworks.

Changing macroeconomic environment combined with social change

A changing macroeconomic is mainly as a result of changing global economic conditions. As a consequence, the economic power of governments and their ability to influence markets has been considerably reduced due to the rise of multinational and transnational companies. Governments' therefore have found it increasingly difficult to intervene directly in markets, and hence the government's role in certain markets has changed more towards that of one of a facilitator. Reduced economic power has also meant that governments can no longer afford to simply give public transport operators an open budget with which to provide services. Over this time there has also been major social change, with changing mobility patterns and overall increases in levels of personal travel. In particular, less people are now reliant on public transport services, thus the gap between costs and revenues have grown substantially causing an increased subsidy requirement. One 'solution' to this was to cut transport services, a view best encompassed by the 'Beeching' approach, where unprofitable railway lines were closed in the 1960s in the UK in search of the profitable railway. An alternative approach is to attempt to run the existing network but at a reduced cost, hence the motivation for reform.

The desire to introduce competition into the provision of transport services

This is partly connected to increasing discontentment with public service provision. The Morrisonian view of transport provision is that of a public corporation being the sole

provider of transport services. Apart from the x-inefficiency argument cited above, reform may be motivated because of social change and the increased desire for choice and options. Whilst fifty years ago people would generally put up with the bus service provided by the local city corporation regardless of the quality of the service, now they won't. Choice is more a part of today's society than it has been in the past, hence introducing competition into the market gives the consumer more viable choices.

OWNERSHIP FORMS IN THE PROVISION OF TRANSPORT SERVICES

We end the unit with a very brief look at ownership forms in the provision of transport services. It may come as a surprise that even where considerable market reforms have been enacted in the provision of transport services, there still remains a wide range of ownership forms within the sector: public ownership is far from being completely removed from the provision of transport services. Ownership structures range from a tightly controlled government department to a free market profit maximising company. In many ways the issue of the ownership of transport assets is very closely associated with the issue of transport governance, as those empowered by the state to provide transport services are generally those that own or control transport assets.

Figure 5.4 attempts to give, in brief summary, an overview of the main ownership forms, public and private, with the focus on the economic rather than the legal issues. Some cross over between the two however is inevitable.

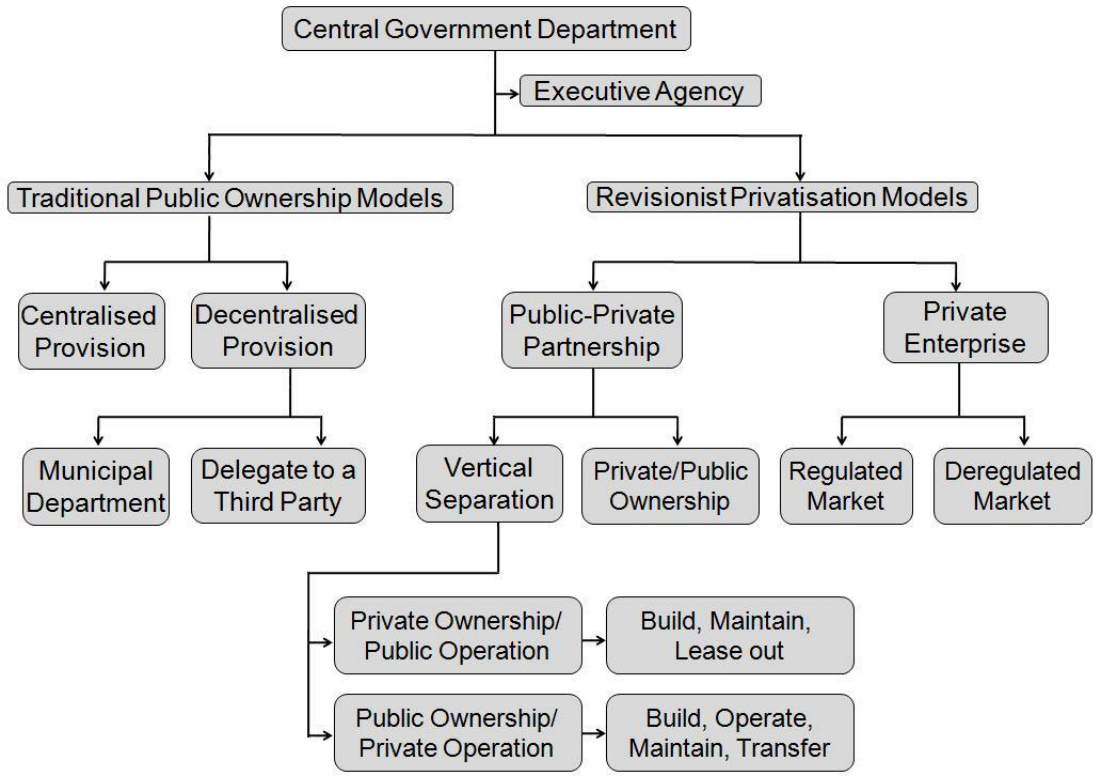


Figure 5.4: Ownership forms in the provision of transport services

Source: Cowie (2010).

Space prevents going into detail what each of these are, however most should be obvious to the intelligent reader! More detail however is given in Cowie (2010).

SUMMARY AND REFLECTION

This unit has examined the issues of regulation and ownership in transport organisation and operation. As should have been seen in the course of the text, there is a balance somewhere between the regulation of transport markets through the mechanisms of ownership and regulation. The old style model is encompassed in the ideas of what became known more generally as the Morrisonian model, where a national public corporation is a publicly owned company that operates at arm's length to the state with a remit to act in the public interest and break even. Such a model originated from a transport problem, i.e. London's public transport. Various drawbacks however were identified and encountered with this approach in public ownership, and much transport provision across the globe has been privatised or reformed. Much however still remains in public ownership, with buses normally privatised but metros, trams and rail services in most cases still remaining in public ownership. The problem now has switched from trying to overcome the problems with public ownership to trying to steer the private sector towards doing what the public authorities want it to do through the regulatory mechanism.

CHAPTER EXERCISES

Exercise 5.1 – Regulating the Price

In the following exercise, you are given some basic information for a local bus operator for the first year of operation and then asked to find the regulated average price over the following two years. In the following calculations, you should round all figures to the nearest whole unit (i.e. nearest £000s, 1000 journeys or the nearest pence in terms of the actual price). Information on the first year of operation is thus given below:

	Year One
Operating Costs (£000s):	500

Journeys in thousands:	600
Profit Margin:	16%
Average Fare Charged:	0.97
Revenue (£000s):	580
Profit (£000s):	80

As seen above, the level of profit margin in the first year is 16% at an average fare of £0.97. You are given the following information below in order to set the regulatory price:

Expected inflation in prices and operating costs in each of the next two years:	3%
Rise in passenger traffic in each of the next two years	1%
Expected annual efficiency improvements:	2%

Increases in passenger numbers are expected to arise from other external factors that impact on the local bus market, therefore you should assume that this growth is not dependent upon the final price that will be set. It is also anticipated that this increase will be incorporated on existing available capacity, thus any changes in passenger numbers will have no effect on costs, therefore any increases in costs are purely as a result of inflation.

a. You should calculate the regulated average price (to the nearest pence) for the following two years on the following two basis:

   i. On a straight RPI-X% basis where X is set at 1%
   ii. On a return on capital employed basis, where the regulatory asset base is set at £1.2m and the rate of return at 8%.

b. Now re-calculate both a(i) and a(ii) on the basis of an anticipated rise in passenger traffic of 2% over the next two years.

c. Using your answers to parts a and b, outline the advantages and disadvantages of regulating based upon RPI-X% as opposed to a return on capital basis in an expanding and a declining market.

d. The company proposes to undertake an increased investment programme in order to enhance existing bus services. It estimates that it will invest £500k up

front and write these investments off over a ten year period. It will fund this investment initially from increased borrowings, which on average will be borrowed at a 5% rate of interest over the first two years. It forecasts however that these enhancements will have a major impact on demand, increasing it by 12 per cent in each of the first two years. The authority agrees that £300k of this new investment can be added to the regulatory asset base. Using your original calculations for part a(ii), what would be the revised regulated price over the first two years that would maintain an 8 per cent return if the proposed investments were made? Should the regulatory agree to this investment plan – note that whilst a 'good' thing as it will increase passenger numbers, are there any other reasons why the regulator should agree to this investment plan?

UNIT SIX
TRANSPORT SUBSIDY

INTRODUCTION

This is the final of the seven units on Transport Economics. After an introduction to Transport Economics, in the first unit, we have looked at the role of individual consumers and their demand for transport (Units 1 and 2) and the factors behind supply and transport costs (Unit 3). The focus remained on public transport operators for Unit 4 and Unit 4b, looking at the market structures under which they operate and their pricing policy. Finally, unit 6 looked at the issue of transport economic regulation and ownership.

In this unit the main focus of the unit will be on the role of transport subsidy. We will look at transport subsidy from the perspective of individual travellers, public transport operators and governmental bodies.

RECOMMENDED READING

The recommended reading for this unit is:

Cowie, J. (2010), Economics of Transport, Chapter 11, Transport Subsidy

And honestly, despite the billions spent by governments throughout the world on transport subsidies, that is the only book that thought the topic important enough to devote a whole chapter to the topic!

LEARNING OUTCOMES

Once you have worked your way through this unit you should be able to:

- Explain the economic rationale behind the payment of transport subsidies
- Explain the difference between supply side and demand side transport subsidies
- Appreciate aspects of transport subsidy such as the rural transport problem and cross-subsidisation.

The Rationale for Public Subsidy

We begin our examination of subsidy by firstly introducing the rationale for public subsidy. In most market based economies, if a firm or industry cannot make a profit from selling its goods or services it goes out of business. Why therefore, should public transport industries (particularly rail) be any different? In order to address this issue, we need to introduce the idea of an externality.

EXTERNALITIES

Externalities are often called 'spill-overs' and they occur when the costs of producing a good or service falls not only on the producer of that good or service but also on others that are not involved in the activity in any capacity. An example would be if you lived next to a motorway. The cost of travel on the motorway would not only fall on the users of the motorway in terms of petrol, travel time, vehicle depreciation and so on, but would also fall upon you in the form of living in an environment with extra pollution and increased noise levels. The full cost of an activity therefore can be divided into two categories. Firstly, private costs, which are those that fall on the producers or users of that product and secondly public costs, which are costs those that fall on others. The full cost, i.e. the addition of private and public costs, is normally referred to as the social cost.

The main implication of externalities is that they are not taken into account when individuals make private production and consumption decisions, as only the private costs of that decision are considered. In terms of an external cost, this leads to over-production of a given good or service, as illustrated in Figure 6.1.

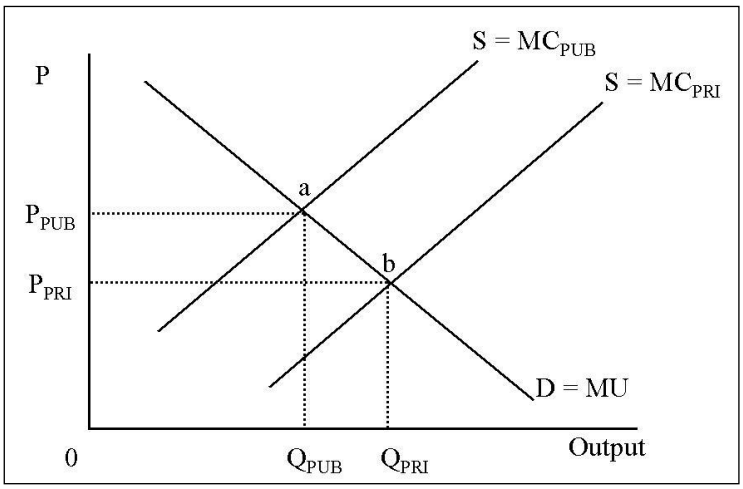


Figure 6.1: Over production due to an external cost

As only private decisions are taken into account in this market, the market would be in equilibrium at point b with a market price of P_{PRI} and quantity traded given by Q_{PRI}. If we bring in externalities however, because these are a cost that are not taken into account,

and costs are a determinant of supply, once these are introduced into the analysis the effect is to shift the supply curve to the left. Hence when all costs are taken into account, the market would be in equilibrium at point a, with a higher market price (P_{PUB}) and lower quantity traded (Q_{PUB}). This is a case of over-production, as not all of the costs associated with this good or service are taken into account in production/consumption decisions.

Externalities however not only work in relation to costs, but there may also be benefits that are not shown in the market place. For example, if an individual chooses to use the train rather than the car for their journey to work, then it is not only that individual who benefits from that decision but also existing road users, as their journey to work time will be reduced. There will also be other benefits in the form of reduced pollution and less accidents (the train is a safer form of transport than the private car). This can be shown in Figure 6.2.

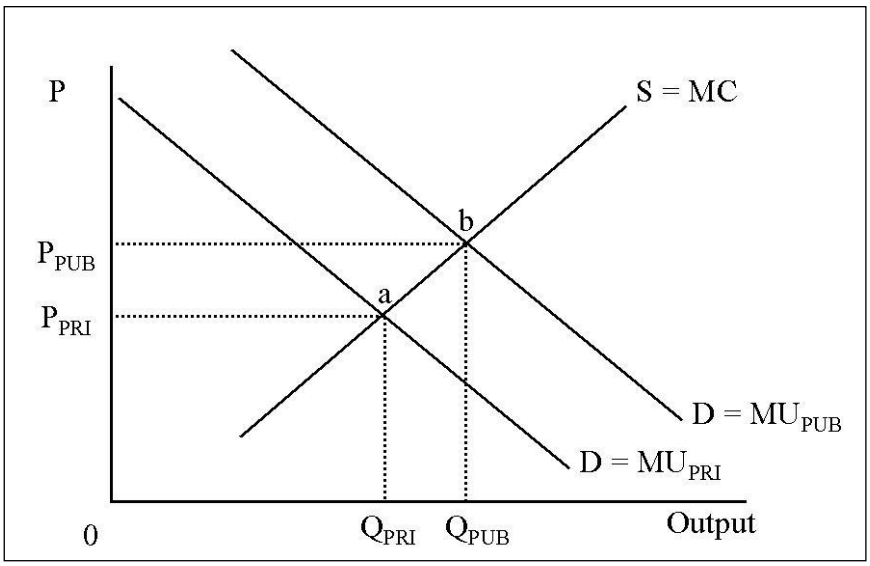


Figure 6.2: Under consumption due to an external benefit

As above, if only private decisions are taken into account in this market, the market would be in equilibrium at point a, with a market price at P_{PRI} and quantity traded given by Q_{PRI}. If we bring in externalities however, because in this example these are a benefit that are not taken into account, this would have the effect of increasing demand at each and every price. This is because at each and every price there is a higher level of utility (benefit) than is being registered in the market place, as only private benefits are valued. Once these are taken into account, the effect is to shift the demand curve to the right. Hence when all benefits, both private and public, are taken into account the market price should be higher (P_{PUB}) and the quantity traded higher (Q_{PUB}). This is therefore a case of under consumption, as not all of the benefits associated with this good or service are taken into account in the production/consumption decision.

In both of these cases, over production and under consumption, we have what is known as market failure – the market as such does not find the 'right' answer in terms of levels of production and consumption. By taking this second example further, we can introduce our first rationale for subsidy, and that is to correct for market failures. If we concentrate on the quantity traded position, then the government could pay a subsidy to operators to produce the same effect. This is shown in Figure 6.3.

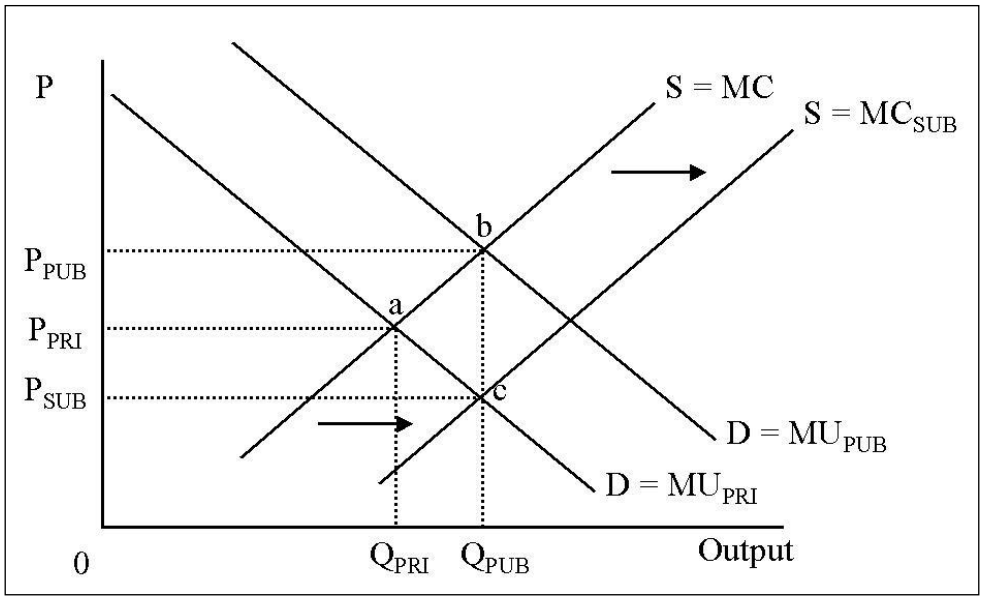


Figure 6.3: Subsidy to operators to correct for under-consumption

Without the payment of subsidy, the market would be in equilibrium based solely upon private decisions, hence this would be point a, with a quantity traded shown by Q_{PRI}. We have already identified above that this level of quantity represents under consumption. In order to correct this market failure, the government could pay a subsidy to operators. The effect of paying this subsidy would be to in effect lower operators' costs. This would cause a shift in the supply curve to the right, hence more is supplied at each and every price. The market would be in equilibrium at point c, with a quantity traded of Q_{PUB}. The difference in the prices between P_{PUB} and P_{SUB} is the effect of the subsidy. Note that in this example, although we identified this as a demand side problem (under consumption), the easiest 'solution' is a supply side measure (pay a subsidy to operators).

Correction of such market failures is undoubtedly the strongest rationale for the payment of a subsidy, and also moves away from the idea that subsidy is paid simply because these firms are massive loss makers (as was often the case in the 1970s with 'key' employers). Without the payment of subsidy in such cases, these transport services would simply not be produced in the same quantities as they are, and the costs of such a

situation in terms of added pollution, road congestion and safety would considerably outweigh the actually subsidy paid. This however is not the only rationale for the payment of subsidy, and the two other main ones are:

1. Role of transport in economic development. This is something you will have already seen in Transport Appraisal (or will see again in Transport Appraisal depending upon which way round we teach TEA!). Transport can be seen as being the vital infrastructure required to promote industrial growth. This relates to both retaining the integrity of the network and keeping services at a desirable level. No single manufacturer however is going to pay for the upgrade of transport services into such an area, as (a) it probably could not afford to do so and (b) once the services were upgraded it would not be the only firm to benefit from the improvements. Any such developments therefore would only take place with government intervention, either in the form of capital grants for infrastructure projects or the payment of subsidy for service upgrades.

2. A range of social considerations that can be grouped under the heading of equity. Whatever the level of car ownership, there are groups who are unable to be car drivers. This particularly refers to the young, the elderly, various groups of disabled people and people unable to afford a motor car. These people tend to be on low incomes and may not be able to pay the full costs of the public transport they require. This is particularly true as growing car ownership increases the cost burden on remaining public transport passengers. There are also social inclusion grounds for providing transport services on the basis of a 'social necessity'.

The second reason for subsidy, on grounds of equity, has an underlying assumption that transport services, or at least some of them, should be allocated according to "need". Transport services are usually allocated according to demand, a basic premise that underlies much of the module. It can be argued that transport is a need, because a civilised society should allow its members to be able to access their work, shops, recreation and other activity locations on a 'fair' and equal basis.

TRANSPORT SUBSIDY

Today transport subsidies increasingly come in the form of a contractual payment for a service and may be applied to a specific good, or to a range of goods, or to the whole activities of a producing firm. Subsidy is used as a lever in the transport industry to hold down public transport fares and maintain public transport services.

Levels of transport subsidy

Subsidy levels are affected by the efficiency of public transport operators, the competitive environment, the pricing policy, the size and nature of the public transport system and government policy through the evaluation of subsidies. We have looked at many of these

factors in the earlier parts of the module. If these factors were to remain constant and it were possible to ensure that any increase in subsidy funds did not leak into unintended purposes (which may be unrealistic), then those extra funds can be used to enhance the public transport service. Such enhancements would include factors such as higher investment, lower passenger fares, lower freight fares, more frequent vehicles, a better quality service or any combination of these giving public transport an optimistic future.

Transport subsidy complicates the pricing problem, because if central or local government provides the subsidy for a specific service, then it may be seen as representing that government's demand for that service and treated alongside the demand of other customers. When lump sum subsidies are given to transport undertakings for general revenue purposes, however, problems arise in deciding upon the best methods of using the subsidy and the appropriate charge to levy on customers. The technique used to compare the return on this form of expenditure with other state expenditure is called cost benefit analysis.

Transport subsidies used to buy entire operations in the UK, particularly before privatisation of the railways and deregulation of the bus industry. However, problems stemmed from the fact the subsidies were open-ended and there was some evidence that increased levels of subsidy resulted in increased operating costs. In other words, service levels were not benefiting from the subsidy. Hence, through the deregulation of buses and privatisation of the railways in the UK, subsidies became much more targeted, on routes or even parts of routes. The aim was to obtain value for money public transport subsidy. However, as the franchising process proceeded during rail privatisation, bids from the train companies showed an increasing tendency to promise ever more ambitious reductions in subsidy.

We know using evidence from around Europe that there are cost efficiency gains (after allowing for administrative and monitoring expenses) from competitive tendering of between 10 and 30% using competitive tendering over more general subsidy arrangements[19].

Table 6.1 shows historical levels of urban transport subsidies in the UK before, during and after deregulation and privatisation (data pre-1992), in order to illustrate the considerable rise in subsidy levels that occurred in the 1970s and 1980s – prior to this period, subsidy for transport services although an issue, e.g. Beeching!, had never been a major problem. After the 1968 Transport Act, explicit powers for subsidies were given to local authorities. Changing economic conditions, increased mobility and a significant modal shift from bus to car meant that transport subsidies rose tenfold during the 1970s, with a particularly dramatic rise during the early years of the decade. Some subsidy growth was due to cost increases, as shown in the table by figures adjusted for inflation. Under successive Conservative governments in the 1980s, much more emphasis was placed on market-determined provision with the object of reducing subsidy levels. Hence

[19] Cowie (2010), *Economics of Transport*.

the downturn in transport subsidy, although it was still higher than 1970s levels. For more up-to-date data, you should consult the DfT's excellent annual publication 'Transport Statistics Great Britain', which is available freely from their website. This will give accurate data on the bus industry, but is sketchy with regard to the railways. As has been noted earlier in these course notes however, the railways receive around £4Bn a year to run the railways, hence that total does not include investment and upgrades to the network.

It should also be noted however that transport subsidies, particularly those relating to rail, have in the past been inversely related to economic activity. When economic activity is high, as was the case in the late 1980s with the 'Lawson Boom', this produces a higher demand for rail services (remember, demand for transport is derived), and hence more passenger revenue is generated through the fare box. The requirement for subsidy is therefore reduced, although more recent trends have completely bucked this trend, with rising numbers and rising subsidy levels, and the economic downturn of the late 00s is unlikely to have the same impact on subsidy levels than would have been the case in the past. Research presented by Rye (2009) for example, suggests that the link between GDP and rail travel has reduced considerably since 1995. This however is a whole other issue, and space and time do not allow us to explore this issue further. It is however connected to the changing economics of public transport provision, and one that you may examine further later on in your career.

Table 6.1. Subsidies and grants for UK bus and rail services (£m)[20]

	1971	**1974**	**1979/80**	**1984/5**	**1987/8[a]**	**1989/90[a]**
Rail[b]						
Local Government	5	18	114	211	78	104
Central Government	79	394	643	956	106	819
Bus						
Local Government	15	122	410	902	673[c]	668[c]
Central Government	31	55	133	128	150	174
Total	**130**	**589**	**1300**	**2197**	**1907**	**1765**
at 1985 prices	**613**	**2089**	**2046**	**2281**	**1745**	**1572**

[20] Else PK (1992), *Criteria for local transport subsidies*, Transport Reviews 12, p293.

Totals at 2008 prices	1180	3997	3586	4229	3282	2586
Of Which Bus:	762	2796	2088	2246	317	1352

[a] Rail subsidies in 1987/8 and 1989/90 reflect transfer of responsibility for London Transport from GLC to Central Government.

[b] Includes London Transport rail services.

[c] These figures include revenue support for bus services in London, despite the transfer of responsibility to Central Government.

It is worth stressing however, that privatisation of the rail industry in Great Britain has not led to reductions in subsidy. The figure of around £4.0bn paid in subsidy to the rail industry is far higher in real terms (by a factor of around 3 in real terms) than was ever paid in the past. The reasons behind why this is the case however, is out with the scope of this module and is covered in the Public Transport module.

Cross-subsidisation

Privatisation and deregulation have resulted in public transport markets being often unstable, with companies entering and leaving the market relatively quickly. This is not good for subsidy. For example, in many towns, large bus companies in the past have used profits they make in other towns where they have a monopoly to subsidise their bus fares and thereby drive competitors out of business, only then to raise prices above those that the competitors had been charging. This is known as predatory pricing[21], and is another aspect of price discrimination, not mentioned in the previous unit.

We can define cross-subsidisation between routes or activities as occurring when one activity does not meet its full costs and is supported by another more profitable activity. This cross-subsidisation can exist at many levels, a bus service may even be cross-subsidised by a profitable operation on the same route that runs at a different time of day and/or day of week. By far the biggest argument against cross-subsidisation however, particularly from an economist's point of view, is that it hides the true costs of providing a service. If profitable services are used to support loss making routes, then the true costs of providing these services is completely hidden. In turn, if the true costs are not known

[21] It should be clarified that predatory pricing relates to the situation where an operator uses profits from one market to cover losses in another market with the clear intention of driving the competition out of business. If the firm simply charges a lower price because it has lower costs, which then drives rivals out of business, this is not predatory pricing but rather fair competition. Predatory pricing is anti-competitive and illegal (it is what American's call 'anti-trust').

then there is no way of judging whether these are worth supporting or if the money would be better spent on other transport services. The second key point is that it is actually those on the profit making routes that are subsidising the loss making services, not the operator's profits. There are far better mechanisms that can be used, such as a direct payment, to ensure that socially necessary services are provided. That way everyone, at least every tax payer, contributes to paying for the service and not just those on profit making routes. These are some of the reasons why there has been a significant move away from large block subsidy grant payments to more route and network specific subsidy payments.

INTERVENTION IN THE MARKET

In order to pay subsidy to correct for the market failures outlined above, subsidies from transport authorities can either be 'injected' into the supply side of the market or the demand side, known conveniently as supply side and demand side measures.

Supply Side Measures

A supply side measure is where the subsidy is paid directly to the operator in order to 'allow' the operator to supply to the market a level of services that it would not have been able to do so otherwise. Services that are provided as a consequence of supply side measures are open to all and not specifically targeted at certain groups within society. The direct effect therefore is to increase the supply of that service to the market. This is illustrated in Figure 6.4, which is a minor, put important, amendment of Figure 6.3 outlined earlier. In that diagram the market is in equilibrium at point a with a market price of P_{PRI} and a quantity traded of Q_{PRI}. As was illustrated, due to an external benefit however the 'true' market value, i.e. when all cost and benefits are taken into account, should have been at point b with a quantity traded of Q_{PUB}. In order to correct for this market failure, one option for the transport authority or state is to increase the level of supply so that the market will be in equilibrium with a quantity level at Q_{PUB}. Using a supply side measure, this it does through the payment of a subsidy to operators, hence producing the result shown below in Figure 6.4.

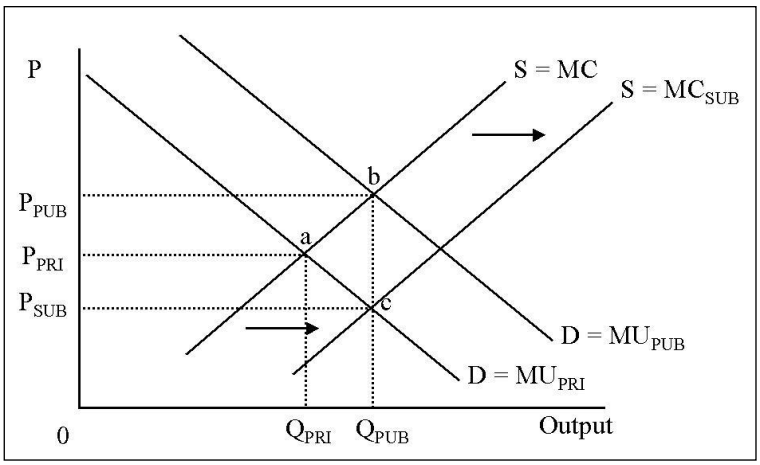


Figure 6.3: Subsidy to operators to correct for under-consumption

Paying a subsidy to operators causes a shift in the supply curve to the right, hence more is supplied at each and every price. The market would be in equilibrium at point c, with a quantity traded of Q_{PUB}. The difference in the prices between P_{PUB} and P_{SUB} is the effect of the subsidy, with the state effectively paying the operators the value of the external benefit. Note also that the extent to which subsidy impacts upon either reducing fares or increasing the level supplied will be dependent upon the elasticity of demand. Where demand is inelastic, the effect of the subsidy will be to mainly reduce the price with little impact upon the quantity traded, whilst in markets with elastic demand most of the impact will be on the level supplied with little effect on price. This does have major implications on transport markets, where as was seen in unit 2 demand for most transport services tends to be inelastic. This is one of the reasons why it is very difficult in general for transport planners to produce a modal shift from the car to public transport services using subsidies alone. Note also that in this example that the external benefit was identified as a demand side problem (under consumption), however the easiest 'solution' in most cases is a supply side measure (pay a subsidy to operators).

Demand Side Measures

On the other side of supply side measures are demand side measures. In theory these are far more straight forward, as a demand side measure is used to correct for a demand side market failure and exist where specific groups or individuals are 'targeted' to receive the subsidy, hence are not open to all. In effect therefore, the individual is given a 'concession' (a reduced fare) to use transport services, either public or private, but in reality most if not all subsidies that are demand side concern some concession in the use of public transport. This usually requires some proof of entitlement to the reduction and should not be confused with commercial initiatives taken by operators such as the Young Persons Railcard which are taken for purely commercial purpose. Concessionary fares are given to individuals that the state has decided should receive some form of discount on their travel needs. These are normally taken on social inclusion ground principles and are specified at the minimum level.

In Britain, the criteria for re-imbursement of operators for the carriage of concessionary passengers is that the operator should be no better and no worse off as a result. How this is done is illustrated in Figure 6.5.

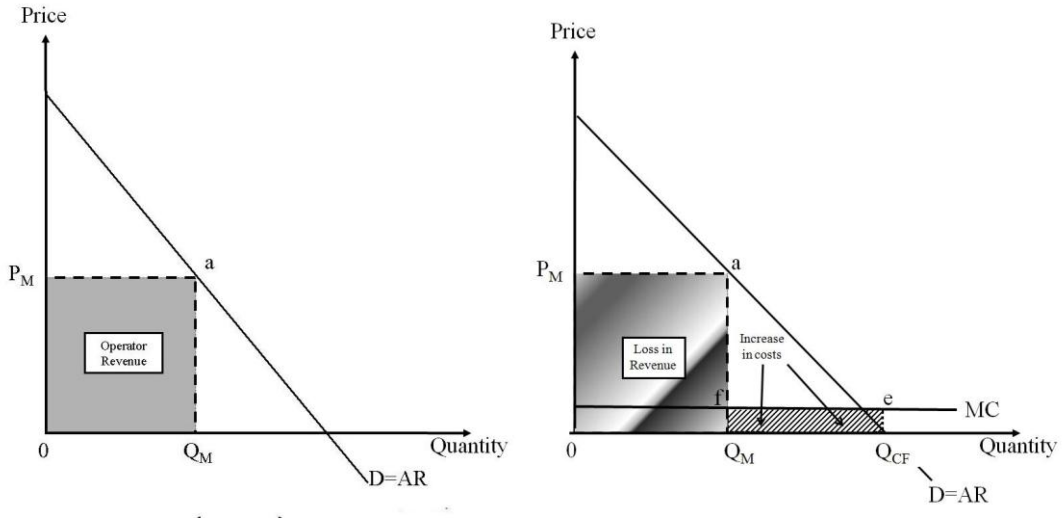


Figure 6.5: Concessionary Fare Reimbursement

On the left is the position for the concession travellers before the introduction of the scheme. Hence they pay the market price of P_M and Q_M demand bus services. All of that revenue accrues to the operator. With the introduction of a free concession however, then all of that revenue is lost, hence the first part of compensation to the operator is the revenue lost as a result of the scheme. In order to be left 'no better or worse off' however, that is not the only reimbursement that is required, as with the free fare demand will increase from Q_M to Q_{CF}. This is known as the generated traffic. There will be an added cost in carrying these passengers, and as most are in the off peak periods this will be compensated at the marginal cost. How this generation factor is calculated is by using an estimate for demand elasticity, hence if we know how many concessionary travellers there are in total, then by applying an elasticity of demand we can work out what the generation factor will be. You should think through this process, although one of the exercises includes a simplified version of operator re-imbursement.

DRAWBACKS AND METHODS OF PAYING SUBSIDY

As highlighted in the introduction, there are many more issues that surround the actual payment of subsidy than subsidy itself. Most, if not all, agree it is a 'good thing', however most if not all recognise that there are many issues that need to be considered first and there are several drawbacks from an economic point of view where subsidy is paid. The main drawback of paying subsidy, to which many of the other issues can be related, is that it is always a second best solution. The best solution as such is always provision of the service by the market under the conditions of perfect competition. Why this is the case is that in the instance of using subsidy, then this normally requires some form of regulation. With regulation comes added bureaucracy and the need to measure and

monitor performance and associated added costs. It is for this reason that it is a second best option. It also means that the payment of a subsidy interferes with the market signals generated by the forces of demand and supply. In a free market situation (and again assuming perfect competition), then the market will ensure that the 'right' services are provided in the 'right' quantities. If consumers desire more of one type of services over another, then demand for the former will rise and for the latter fall. This will be signalled to producers through the price mechanism and they will ensure that more of the former service is provided. When subsidy is paid however this interferes with the price mechanism and the whole process is lost and becomes dependent upon accurate forecasting and planning.

Methods of Payment of Subsidy

Today transport subsidies increasingly come in the form of a contractual payment for a service and may be applied to a specific route or a batch of routes or network. Where these are paid to a private sector operator, this would normally be in the expectation that the operator could provide the service at a lower cost than a publicly owned firm and that cost savings would more than offset the operator's profit from providing the service. The main types of subsidy payments are:

Deficit Subsidy

This is the simplest form of subsidy and as suggested by the name, this is where the authority pays the difference between the revenue received from the service and the cost of providing that service. In the past these have been in the form of open-ended subsidies, where the authority covers the size of the deficit irrespective of how large that subsidy is. Increasingly however, authorities have moved to a form of an allocated or negotiated budget. In many respects these forms of subsidy are consistent with the traditional view of the nationalised operator where the operator is allocated a budget, usually on an annual basis, to provide public transport services in the authorities' area.

Net Cost Contract

Under such agreements, the operator in effect acts as a sub-contractor to government to provide transport services within a given area. This also reduces cross-subsidisation and produces greater visibility as to the actual cost of providing a given service/batch of services. Such contracts are normally for a set period of time, after which point the contract is then re-tendered. This basically is the idea of the contestable market outlined in Unit 4.

Full or Absolute Cost Contract

Under a full or absolute cost contract, subsidy the operator's bid is based on the full cost of running the service or network specified by the transport authority. The fares charged under such a contract are specified by the authority but collected by the operator who then returns the revenue to the authority. The total subsidy paid under such a contract is

therefore the cost of the contract less the revenue raised. All revenue risk therefore rests with the authority and none with the operator. Transport for London for example operate a system of full cost contracts for the provision of bus services.

SELF-ASSESSMENT EXERCISE

Exercise 6.1

Illustrate with the use of graphs the economic argument for the payment of subsidy.

Exercise 6.2

Look over this unit in answering the following questions on transport subsidy:

a. Transport subsidy causes more problems than it solves. Discuss
b. Low subsidy is 'good', high subsidy is 'bad'. Discuss
c. There are more issues surrounding the payment of subsidy than subsidy itself. Discuss.

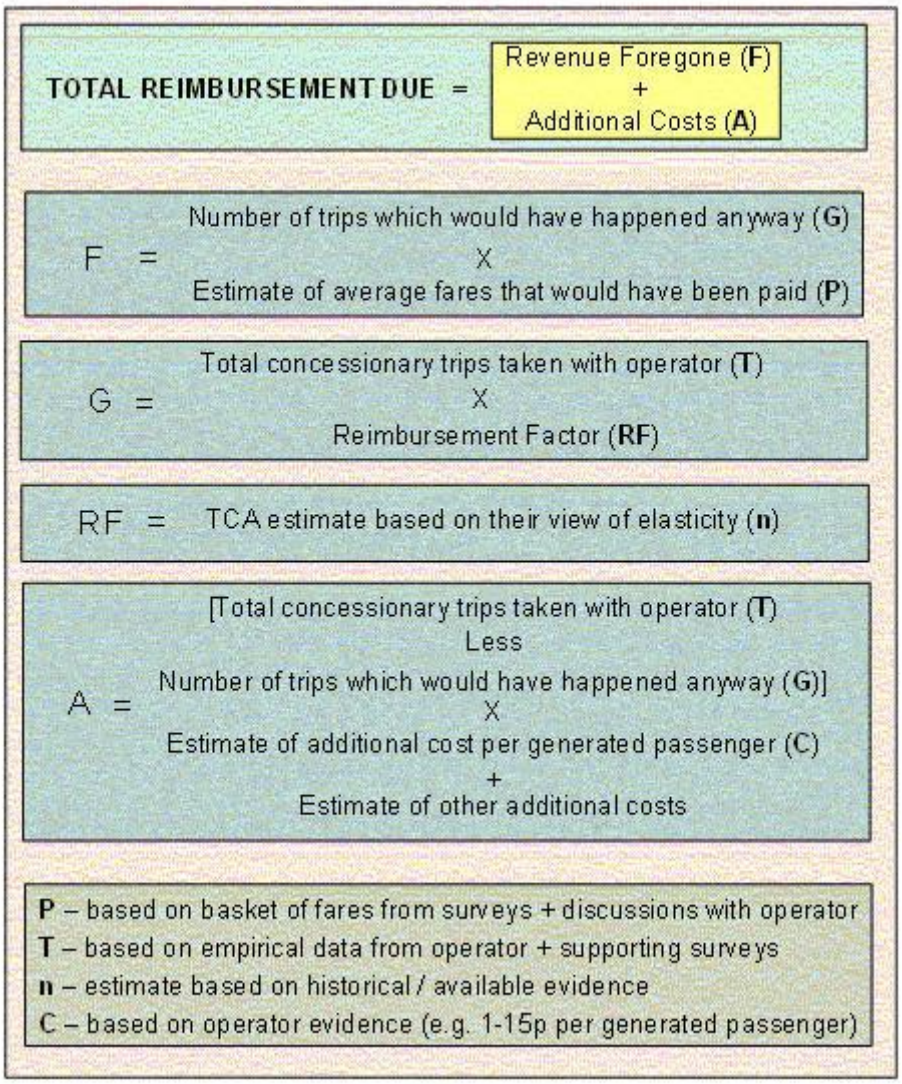
Exercise 6.3

This question concerns concessionary fare reimbursement. Consider the following information:

- Daily number of concessionary trips: 3,500
- Average bus fare: £1.20
- Elasticity of demand: -0.3
- Cost per extra passenger carried: £0.20
- Total number of operating days per year: 360

Using this information, you are required to calculate the annual re-imbursement for the carriage of concessionary travels for this operator. You may find the following diagram, taken from the DfT website, useful:

What information is required?



Source: DfT website, Concessionary fare reimbursement guidelines, information for local authorities and operators, accessed on 23rd November 2009.

The only other piece of information you need to know is that RF is calculated in the following way:

$$RF = \exp(n)$$

Part b – the operator questions whether -0.3 is the right value to use for elasticity, and considers that given the nature of this particular group their demand will be less inelastic. They think a figure of -0.7 is probably more appropriate. Should they suggest using this (i.e. what would the reimbursement be if this revised figure was used?)

Unit 6 Summary

In this unit, the final of the six units on Transport Economics, we have been looking at transport subsidy. As well as understanding the nature of transport subsidy, you should now know the reasons and rationale behind providing transport subsidies and the general levels of subsidy in both the UK and Europe. We also looked in detail at the rural transport problem and the issue of cross-subsidisation.

I hope that you have found the units on the subject of Transport Economics worthwhile and I am sure that you will find the remainder of the Transport Economics and Appraisal module of great interest. One of the major things at this point is that hopefully you see the link between the two subjects (!), and you should not forget the economic component when examining Appraisal.

ANSWERS TO SELF ASSESSMENT EXERCISES

INTRODUCTION

The following pages contain some of the answers to the self assessment exercises in the preceding units. Note that only some of the answers are included. This is not by accident but by design. You should be aware by now that in many instances there is no one 'correct' answer to the given problem, but rather there may be several effects taking place. What is important in these instances, and in any subsequent examination that you may have to sit, is the logic behind any solution. In other words your answer may not agree with others but as long as there is a reasonable logic behind the solution then that is acceptable.

A second reason for not including all of the answers is that you should at this stage in your studies be able to be self critical of your answers. In other words, you should be able to assess the extent to which your responses to your answers are 'correct' or not. If after attempting a self assessment exercise you don't know the extent to which your answer is correct or not then you need to undertake further revision, both of the units and the associated reading.

ANSWERS TO UNIT 1 SELF ASSESSMENT EXERCISES

Exercise 1.1

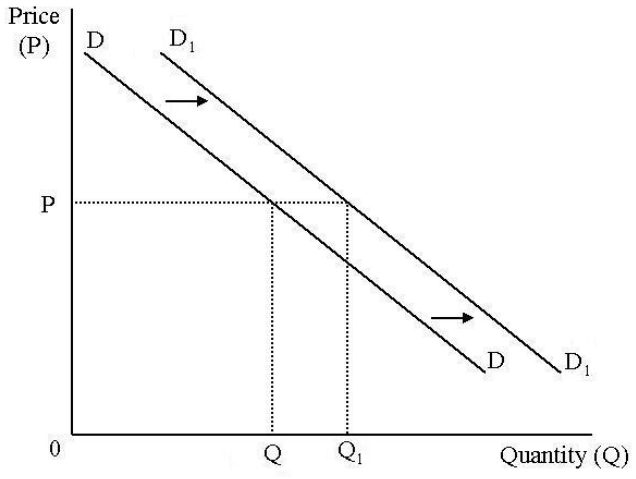
In the text you saw an illustration of the effect of a decrease in the price of bus services (a substitute good) on the demand for rail services. You should use the same process to illustrate what would happen in each of the following cases. In each example you should identify what type of determinant of demand is being examined and your reasoning for your answer (as there may be more than effect taking place).

- A rise in income on the effect of the market for rail services
- A fall in the price of a petrol in relation to the market for new cars
- The effect of issuing travel cards to students that entitles them to a 50% discount on buses on the market for bus services.
- The effect of an announcement that rail fares are to rise by 10% from the 1st of the month following the announcement on the market for rail services

ANSWER

We will begin with the first question, which we will work through in full i.e. if this was under 'exam conditions', this is the answer that would get you full marks. The others we will simply briefly answer.

Rail services are what is termed a normal good, in that as income rises we would expect an increase in demand and each and every price. The demand determinant in this example is therefore income. The effect of an increase in income on the demand for rail services is graphically shown below:

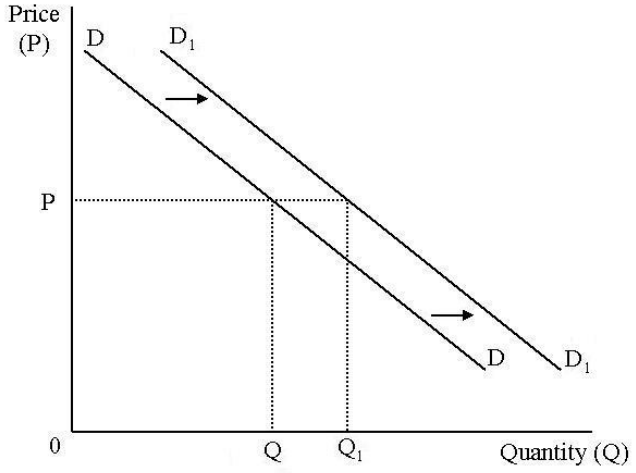


The original demand curve is shown by the line DD, which slopes downwards from left to right. Thus at price P, the quantity demanded would be Q. After the change in income, more rail services are demanded at each and every price. This is therefore represented by a shift in the demand curve to the right. This would be termed an increase in demand, as opposed

to a change in the quantity demanded, which would be illustrated by a movement along the demand curve. Thus at price P, which prior to the increase in income had an associated demand level of Q, would now have an associated demand level of Q_1. Therefore an increase in income would result in an increase in the demand for rail services, other things remaining equal.

A fall in the price of petrol in relation to the demand for new cars

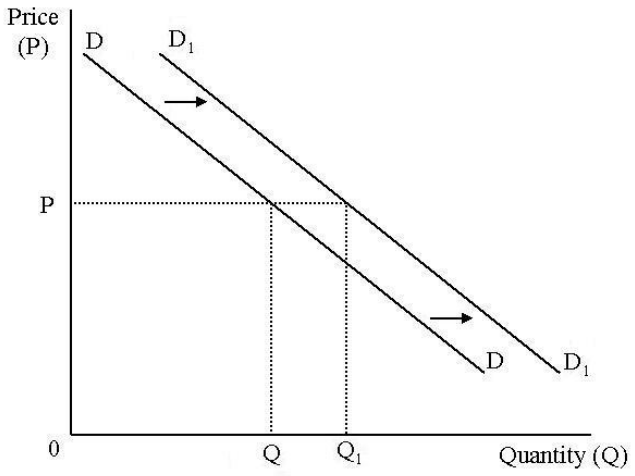
The demand determinant in this example is the price of other goods or services, and specifically, the price of complements. A complementary good is a good or service that is consumed at the same time. Hence bread and low fat spread at complementary goods as they are literally consumed at the same time. What is important in this example is the overall price (cost) of motoring. If the price of petrol was to fall, then we would expect, other things remaining equal, an increase in the demand for all products associated with the activity of motoring, such as lubricating oil, motor insurance, road tax and so on. The example in this case is the demand for new cars, hence we would expect the demand for new cars to increase. This is shown below in the diagram.



The original demand curve is shown by the line DD, which slopes downwards from left to right. Thus at price P, the quantity demanded would be Q. After the decrease in the price of petrol (a complementary good), more new cars are demanded at each and every price. This is therefore represented by a shift in the demand curve to the right. This would be termed an increase in demand, as opposed to a change in the quantity demanded, which would be illustrated by a movement along the demand curve. Thus at price P, which prior to the fall in the price of petrol had an associated demand level of Q, would now have an associated demand level of Q_1. Therefore a fall in the price of petrol would result in an increase in the demand for new motor cars, other things remaining equal. If we wanted to be really really smart in this answer however, we may highlight that other things are most likely not to remain equal. As the demand for petrol is relatively price inelastic, and this is a good that is heavily taxed, then this would result in a fall of the total tax take for central government. Consequently, they may offset the price fall by increasing the duty to re-coup the loss in tax revenue.

The effect of issuing travel cards to students that entitles them to a 50% discount on buses on the market for bus services.

There are actually two possible demand determinants that it could be argued are affecting this example, both of which have the same outcome. The important element is the rationale. Taking the first, it could be argued that the demand determinant is the price of other goods and services, and as above, in this case it is a complementary good. A complementary good is a good or service that is consumed at the same time, and in this case it is the travel card and the rail journey. After the issue of the free rail cards, then more people (students) are likely to travel on the railways, as the overall cost of rail travel (for this group) has fallen. Thus we would expect a higher level of demand at each and every price, as shown in the diagram below.



The original demand curve is shown by the line DD, which slopes downwards from left to right. Thus at price P, the quantity demanded would be Q. After the issuing of the rail cards, more rail travel is demanded at each and every price. This is therefore represented by a shift in the demand curve to the right. Thus at price P, which prior to the issue had an associated demand level of Q, would now have an associated demand level of Q_1. The overall results would be an increase in the demand for rail services.

An alternative explanation could be taste or fashions etc. As the rail card is free (i.e. a complimentary good!), it could be argued that this would induce a change in people's habits towards rail travel, hence causing a shift in the demand curve.

The example actually occurred in practice, where the Dutch government introduced free rail travel for students. One of the outcomes of this was that a group of entrepreneurial students set up a very successful courier business!

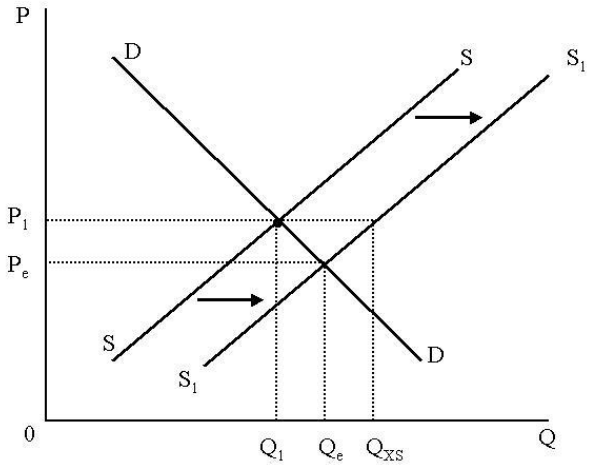
Exercise 1.3

You should now consider each of the following factors on the price and quantity traded for the market highlighted. This time you should identify what side of the market (demand or supply) is being affected, what particular determinant has changed and you should explain your reasoning.

- The entry of a low cost airline into the long haul sector on the market for airline services.
- A rise in income on the market for bus services.
- An increase in fuel duty on the market for bus services
- A fall in the costs of production of bus services on the market for rail services

We will answer the first question in full (i.e. as if under exam conditions), and briefly examine the others.

The entry of a low cost airline into the long haul sector on the market for airline services. This one is straightforward. The side of the market that is being affected is supply, and as it is 'low cost' we would expect a shift in the supply curve to the right. This is because the determinant that is being effected is the costs of production, and these have now reduced (what would happen if it is any old airline?). This is shown in the graph below:



With the reduction in costs, the supply curve shifts to the right from S to S_1 i.e. more is supplied at each and every price. At the existing market price of P_1, there is now excess supply, as shown by $Q_{XS} - Q_1$. This would be represented in the market by empty seats on aircraft. In an attempt to fill these seats, airlines may lower the price i.e. there is downward pressure on the market price. This reduction in price is represented by a shift along the new supply curve. Furthermore, due to this unused capacity, some airlines may leave the market voluntarily due to the lack of sufficient profits and as firms are assumed to be profit maximisors, other markets may represent better profit opportunities. Alternatively, some firms may leave the market involuntary i.e. go out of business. Both of these effects would be shown by a movement along the new supply curve. The effect therefore of the entry of a low

cost airline into the long haul sector on the market for airline services would be to reduce the market price (P_1 to P_e) and increase the quantity traded (Q_1 to Q_e).

ANSWERS TO UNIT 2 SELF ASSESSMENT EXERCISES

Exercise 2.1

The basic questions to be answered are:

4. Estimate the price elasticity of demand for a price fall from 10p to 8p and for a price rise from 10p and 12p.
5. Was the 10p fare the best fare originally?
6. The company considers lowering the fare to 6p, and estimates that demand will be 8.5 million passenger miles. Nevertheless, in order to incorporate this extra demand the company will to put on extra buses, however. How should it decide what to do?

Taking question one, all we have to do is simply plug these values into our equation for price elasticity of demand. That is:

$$\text{PED} = \frac{\%\Delta D}{\%\Delta P} = \frac{50\%}{-20\%} = -2.5$$

And:

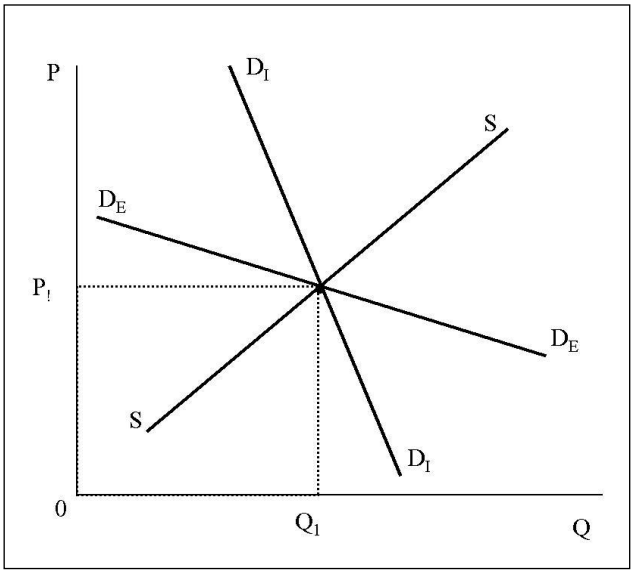
$$\text{PED} = \frac{\%\Delta D}{\%\Delta P} = \frac{-10\%}{20\%} = -0.5$$

The answer to the second question is quite simply no. At the 10p fare, profits were £40,000. If the price had been set at 8p, profits would have been £120,000.

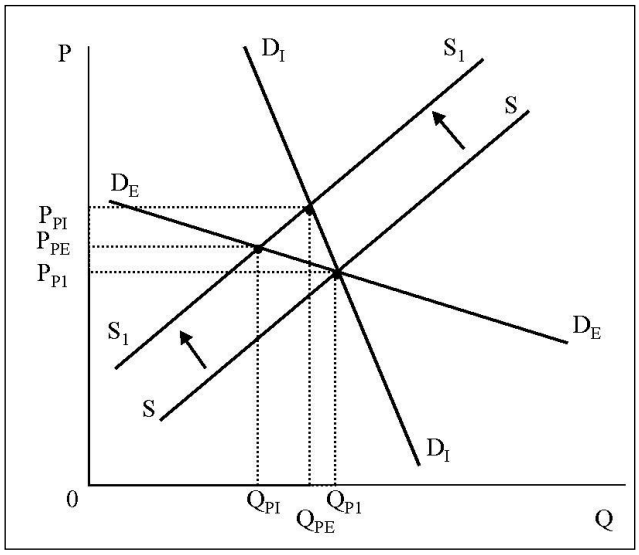
In the third question, the total revenue generated from a 6p fare per mile would be £510,000. This is by far the highest revenue. What it would need to do is calculate the extra cost of providing the extra level of supply to meet this demand. If this is less than £30,000, then this will increase total profits and hence the firm should do it. If however, costs were to rise by more than £30,000 then this would more than offset the increase in revenue and total profit would fall, hence the firm should not do it.

Exercise 2.2.

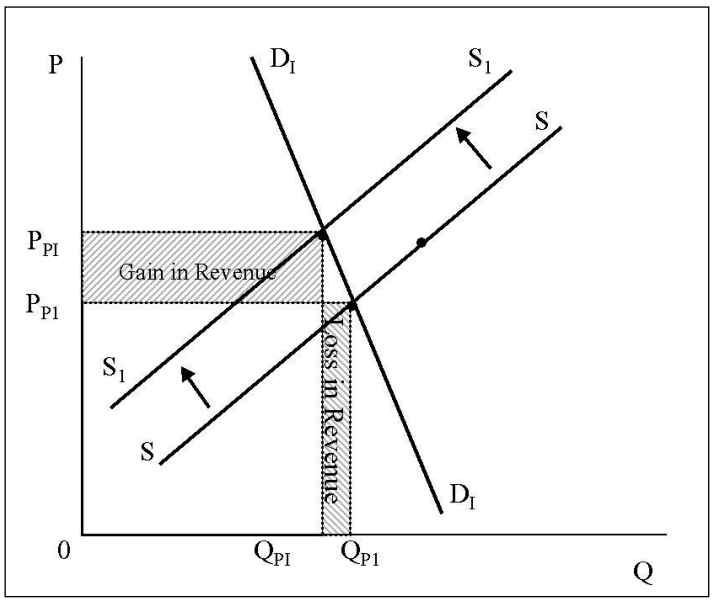
We start to answer this question by drawing out the basic diagram. We know that the effect of increasing a tax will be to increase the price, but we have to show the effect in the market and also the effect on price and quantity demanded pre and post tax increase. The basic diagram to use therefore, is:



The figure above shows the market price at P1 and the quantity traded at Q1. Two demand curves are shown at this market equilibrium, DE which is highly elastic and DI which is highly inelastic. We show the increase in tax by a shift in the supply curve, from S to S1. This is shown below along with the new market price and quantity traded positions for both elastic and inelastic demand:

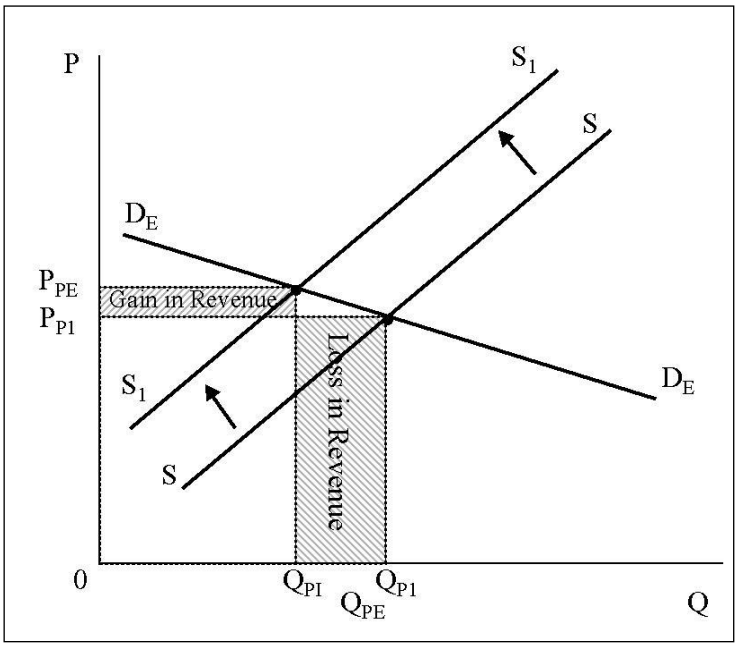


This at first may look slightly complicated, however what we are interested in with regard to tax is the total revenue generate from the price increase. If there is a 20% tax on a good or service, then put simply the tax take will be 20% of total revenue. The tax take therefore will be directly related to total revenue. In order to highlight this therefore, we will actually simplify this and look at each case in turn. Firstly therefore, the situation where the demand is inelastic. This is shown below:



This might be what we might expect through an increase in tax. This increases the price, and the gain in revenue through the higher price charged more than offsets the loss through selling less units. Total revenue therefore increase, and as tax is directly related to the revenue generate, the total tax take therefore increases also.

For an a good with elastic demand, this is highlighted below:



In this case, again the price increases but because demand is elastic, the price increase is far smaller than where demand was inelastic. The quantity traded however has fallen by a far larger amount, hence this more than offset the higher price charged per unit sold. Total revenue therefore has decreased, and as a consequence, despite increasing the tax rate, the total tax take has actually fallen. Note also that most of the increase in the tax in this example has be borne by the industry, whilst in the first case most of the increase in the tax was passed onto the consumer.

The conclusion therefore is that the government should not tax goods that have a high price elasticity of demand, whilst goods with a low price elasticity of demand (i.e.inelastic demand) should be taxed e.g. alcohol, petroleum and tobacco!

ANSWERS TO UNIT 3 SELF ASSESSMENT EXERCISES

Put very simply, the answers are:

Unit	Output (000s)	AP (000s)	MP (000s)	TFC £	TVC £	TC £	ATC £	MC £	AR £	MR £	TR £	Profit £	Price £
(1)	*(2)*	*(3=2/1)*	*(4)*	*(5)*	*(6)*	*(7=5+6)*	*(8=7/2)*	*(9)*	*(10)*	*(11)*	*(12=2*14)*	*(13=12-7)*	*(14)*
0	0	-	-	100000	0	100000	-	-	-	-	0	-100000	-
1	50	50.0	50.0	100000	50000	150000	3.0	1.0	1.5	1.50	75000	-75000	1.5
2	110	55.0	60.0	100000	100000	200000	1.8	0.8	1.4	1.32	154000	-46000	1.4
3	180	60.0	70.0	100000	150000	250000	1.4	0.7	1.3	1.14	234000	-16000	1.3
4	260	65.0	80.0	100000	200000	300000	1.2	0.6	1.2	0.98	312000	12000	1.2
5	350	70.0	90.0	100000	250000	350000	1.0	0.6	1.1	0.81	385000	35000	1.1
6	420	70.0	70.0	100000	300000	400000	1.0	0.7	1.0	0.50	420000	20000	1
7	480	68.6	60.0	100000	350000	450000	0.9	0.8	0.9	0.20	432000	-18000	0.9
8	530	66.3	50.0	100000	400000	500000	0.9	1.0	0.8	-0.16	424000	-76000	0.8
9	570	63.3	40.0	100000	450000	550000	1.0	1.3	0.7	-0.63	399000	-151000	0.7
10	590	59.0	20.0	100000	500000	600000	1.0	2.5	0.6	-2.25	354000	-246000	0.6

How we got there however, is explained on the next page.

OK, let's fully work through one line, and then exactly the same logic should apply for all the other lines. We will begin by actually working through where we have 3 units of the variable factors. Hence, the average product is given by 180/3 = 60, the marginal product is the amount of production added to by the last variable unit employed i.e. the 3rd worker. Hence, it is the difference between the output when 2 workers were employed and when 3 workers are employed, which is simply 180 – 110 = 70. Total fixed costs are easy, these are given as £100,000, and our variable costs are £50,000 for each variable input. Hence for 3 variable inputs it is simply 3 * £50,000 = 150,000.

Average total costs are the total costs divided by the output, to find out how much, on average, each unit cost to produce. Note that our output is given in thousands, hence at 3 variable inputs, the output level is 180,000 and average total costs therefore given by 250,000/180,000 = 1.4 (to the nearest decimal point). Marginal cost is the cost of the last unit produced, and this is where things may begin to get a bit more difficult. As we cannot identify 'the last unit produced', because when we moved from 2 to 3 variable inputs our output rose by 70,000 (and not 1!), then we have to find the average cost of these last 70,000 units that were produced. This is done by dividing the change in costs as a result of moving from 2 to 3 variable inputs with the change in output (i.e. the marginal product) from moving from 2 to 3 variable inputs. Which is (250,000 – 200,000)/70 = 0.7. Thus the last 70,000 units on average cost £0.7 to produce.

This exercise has been taken a bit further as also shown is some revenue figures. Average revenue is simply the price – if for example we sell 100 widgets at £5 each, then the average revenue per widget is our total revenue (£500) divided by the number of widgets sold (100), which comes to £5 i.e. the price. Hence in our example, AR is £1.3. Marginal revenue we will come back to, as we first need to calculate the total revenue. Total revenue is simply the price times the quantity sold, hence £1.3 * 180,000 = £234,000. Finally, profit is our total revenue minus our total costs, hence in this case that is £234,000 – £250,000 = -£16,000 i.e. a loss. Alternatively, if the price was £1.3 and the average cost of each unit produced was £1.4, then we lost 10p on each unit sold. As there were 180,000 units sold, the total loss would be £18,000. Wait a minute, that's a different answer!! Someone in class did raise the issue if rounding to one decimal place would be sensitive enough, and it turns out the answer is no. The actual average cost to two decimal places is £1.39, which would mean that 9p was lost on each unit sold. Hence, 9p on 180,000 units gives an answer of £16,200 (the 200 difference is due to rounding).

That just leaves the marginal revenue. Technically, the marginal revenue would be the revenue received for the last unit sold. However, the law of demand is defined as "number of units per unit of time that consumers purchase at any given price" (page 8 of the course notes) i.e. per time period, there is only one price (this is an assumption we will relax later). Hence, we do not for example sell 6 units at £5 each and then sell the seventh at £4 (hence marginal revenue would be £4), but rather in order to

sell one extra unit we have to reduce the price of ALL units sold (basic law of demand, in order to sell more we need to cut the price of all units). What marginal revenue is therefore is the additional (or reduction in) revenue generated by selling one extra unit (approx!). Therefore, at a price of £1.4 (where we have two variable factors), total revenue is £154,000. By employing a third variable factor we can increase output to 180,000 BUT we need to cut the price of ALL units in order to sell all the output. Hence the price is set at £1.3, and total revenue is £234,000. The change in revenue is positive, and has risen by £234,000 - £154,000 = £80,000, output has risen by 70,000 (i.e. the marginal product), therefore the marginal revenue is 80,000/70,000 = £1.14. Note that marginal revenue at each level of output is always less than average revenue i.e. in order to sell more, we have to cut the price.

UNIT 3 - ANSWERS TO CASE STUDY.

1. The traditional view of railway economics is that the infrastructure and operations are an integral part of the railway system. Economies of scale are associated with 'the railway' as such and are considerable. Company size therefore needs to be very large in order to capture all economies of scale i.e. produce at or near to the minimum efficiency scale level. Given the market size however, it may well be the case there is therefore only enough 'room' for a single operator.

The revisionist view is that economies of scale are only associated with the infrastructure not with service. Therefore, whilst the size of the infrastructure company is important and should be retained as a single organisation (in order to achieve all economies of scale), the size of service companies is unimportant (as the belief is that there are no scale effects associated with this activity).

2. The nub of the answer to this question is the relatively high level of capital inputs used in the production of rail services. In fact, the rail industry is said to be a capital intensive industry (employs a high level of capital). As with most capital intensive industries, the more that the equipment can be worked in the production of output then the lower average costs will be – idle time is money! Whilst you would not employ a worker to sit around all night doing nothing, many firms do exactly that with capital inputs. Whilst accepting it is not quite the same thing (our machine will not be knackered in the morning having spent the night reading the paper whilst our worker will be!), it does put forward the basic idea. Larger railway companies are better able to employ their capital inputs than smaller railway companies. Large parts of the infrastructure in particular, are more likely to be utilised (hence the revisionist argument). Other sources associated with capital are specialisation of equipment and the carrying of spares and maintenance (returns to scale), as well as bulk orders, investment and R & D and financial economies (economies of scale).

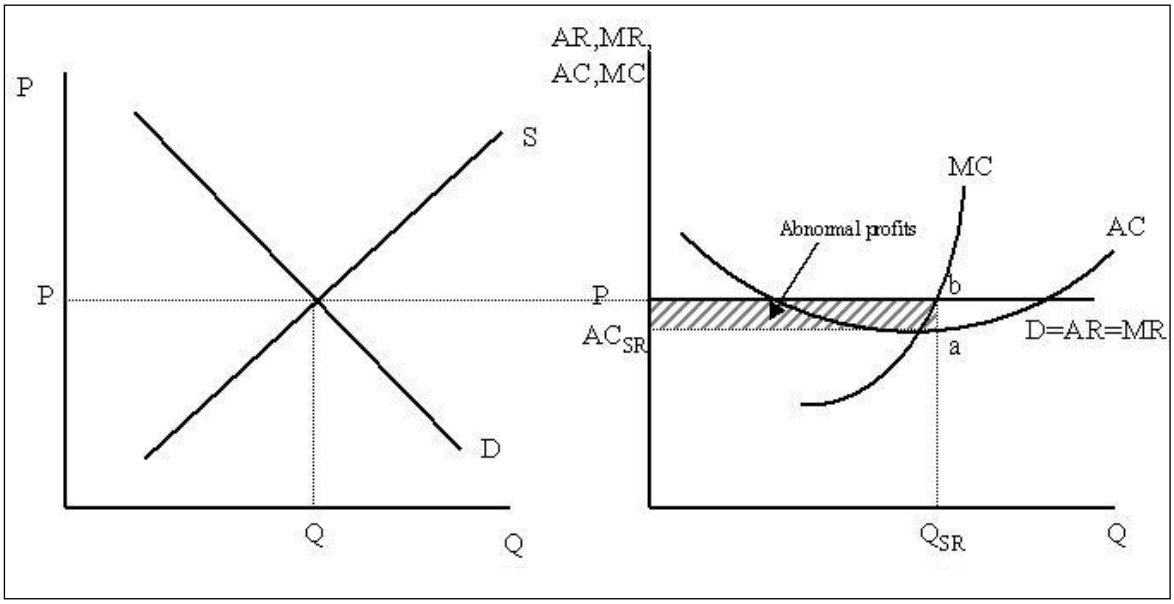
3. Three key factors – bureaucracy, loss of the sense of the individual and X-inefficiency. The Italian State Railway was a classic case of all of these features – many tiers of management (bureaucracy), massively over staffed normally recruited on dubious recruitment practices (loss of the sense of the individual) and a general management slackness (X-inefficiency).

4. The case study highlights that the evidence suggests around 4 or 5 operators. This is purely on the basis of minimising average costs i.e. 'capturing' all economies of scale. Three other main issues however would need to be considered. Firstly, competition, or more accurately, potential competition. If we restrict the number of operators then the opportunities for potential competition are far more limited, hence this would suggest that there should be a higher number of train operating companies. Competition would be the 'spur' to reduce x-inefficiency and generally keep costs as low as they possible can be. A second issue is that the railway industry is heavily regulated (which hopefully you should remember from the Public Transport & Terminal Design). One thing a regulator needs is information on performance, but with a smaller number of operators then opportunities for 'benchmarking' become very limited. Hence again this would need to be considered.

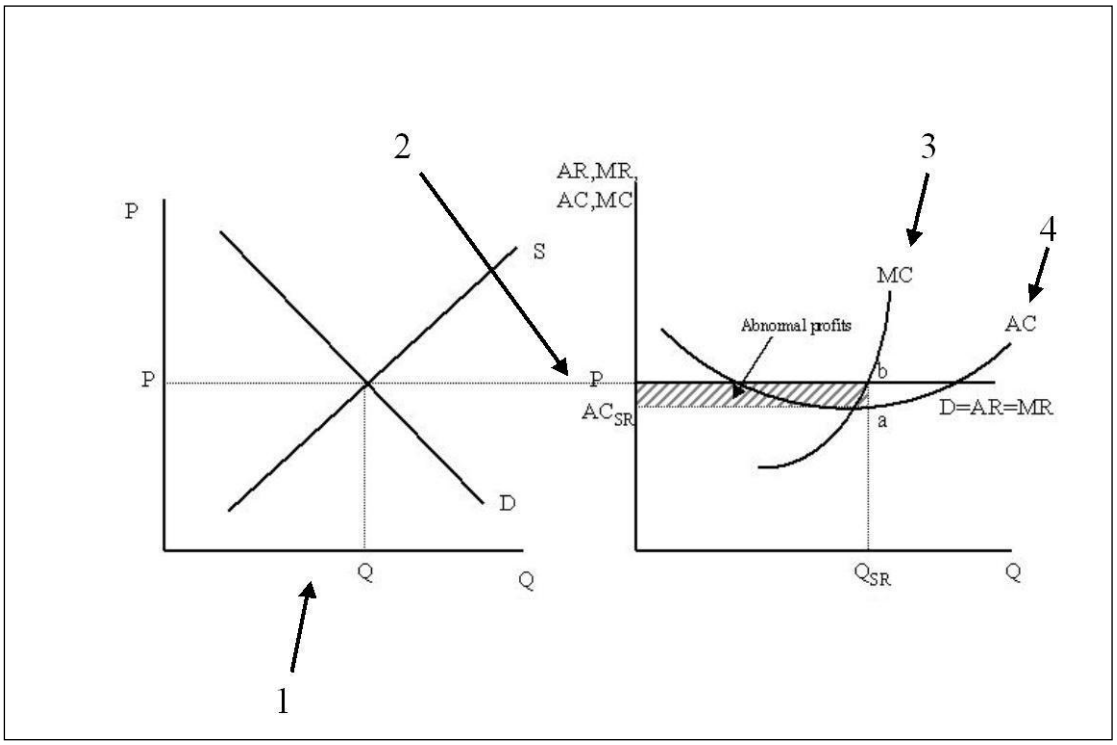
A final point is that there may be some cases where a small, locally run railway, may be a ‘better’ option (both in terms of costs and service quality) than the line being a part of a larger franchise. This is known as ‘microfranchising’ and has never been exploited in the current industry structure.

ANSWERS TO UNIT 4 SELF ASSESSMENT EXERCISES

It is important that you learn to draw these diagrams from scratch – looking at them is one thing, drawing them is another. Your diagram should look something like:



That of course is the end result, what is important here is how we got there. Therefore, refer to the diagram below:

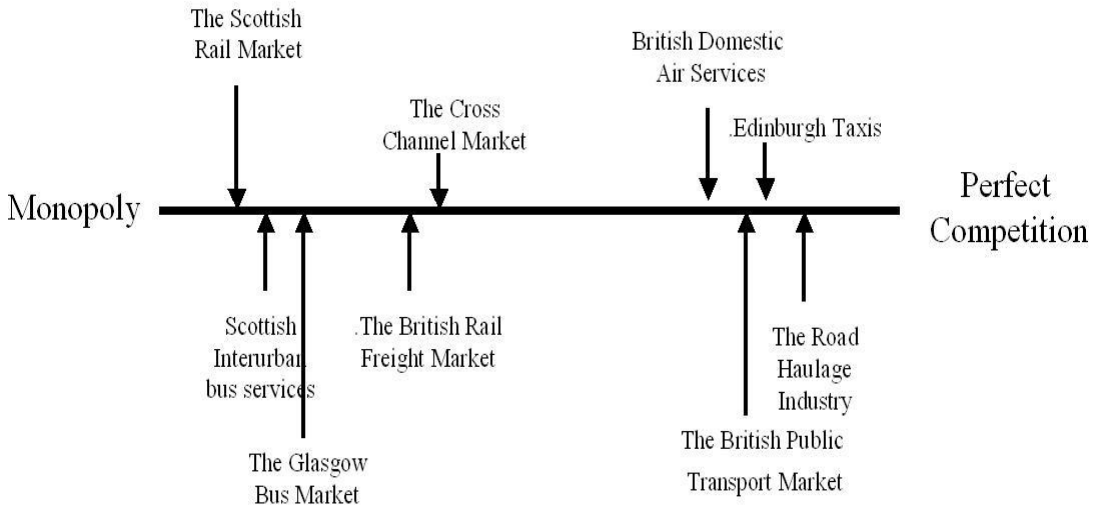


This is the same diagram, but we've added some arrows. How this was drawn was to first draw the market, hence the diagram on the left (1). This gave us the market price, which we then extended across to our diagram on the right (2), to show the position of the individual firm. Under perfect competition, this is also the firm's marginal revenue curve, which we need to use to find the profit maximising position. We then drew the marginal cost curve (3), which gave us the profit maximising position. Once we had that, then we could show any situation in this market, specifically either abnormal profits, normal profits or even losses. This is because the last element we add in this case is the average cost curve, and depending upon where we draw that will give us the result we are attempting to illustrate. If we drew it so that at the profit maximising level of output the average cost was greater than average revenue, this would show a loss. In drawing these, the only thing we have to ensure is that it is the right shape (u-shaped or an inverted parabola as some people find easier to remember(!)) and that the MC curve cuts the AC curve at the latter's lowest point. To show normal profits we would draw the AC curve so that its lowest point was at the intersection of the MC and MR curves i.e. point b above. Finally, to show abnormal profits we would draw the AC curve at a position where at the profit maximising position AC was less than AR, as shown above.

If you are happy with that, now attempt to illustrate the monopoly profit maximising position – in this case, remember that we only need one diagram as the firm's demand curve is the market's demand curve, and draw your curves in the same order as above – hence, marginal cost, marginal revenue, this then gives the profit maximising level of output, then average cost and finally on this one you will have to draw the average revenue. Note that where you draw the last two will determine whether your graph shows an abnormal, normal of loss being made.

There is no substitute for practice with these diagrams, hence apply what you have learned and practice drawing. Simply 'mugging' these up is not recommended under any circumstances, rather you should be able to reproduce them through your understanding and your (practiced) drawing skills. Note by the last comment we simply mean drawing them all in the right place, right shape and that they in some way resemble what a textbook would draw them like.

ANSWER TO EXERCISE ON MARKET STRUCTURES



The Glasgow Bus Market – has one very dominant operator (Firstbus), with limited competition from Stagecoach on some of the main routes.

The Scottish Rail Market – protected monopoly over most of the network – some competition exists, but this is only in a few isolated cases. This situation will continue for the next seven years.

The Road Haulage Industry – mainly consists of small firms offering very similar services, all therefore are price takers. Barriers to entry are relatively low, as all we need is a licence and a lorry. Of all the industries highlighted, this is probably the one that most closely resembles perfect competition.

The Cross Channel Market – only two operators operate the shortest crossing, but a argument could be made that it should be moved to the right if we examine the whole cross channel market (and even further to the right if we bring in air).

British Domestic Air Services – competitive market with reasonably homogenous product and a largish number of operators.

The British Rail Freight Market – only two companies of note and until very recently they offered very different services.

Edinburgh Taxis – homogenous product (its black!) and a degree of perfect information – fares are regulated hence any customer knows they will not (or should not!) be 'ripped off', we also 'know' that they are 'safe', there are a large number of buyers and a relatively large number of sellers. Of all the examples, it's probably the nearest that we have to perfect competition. In some senses however, this is 'artificial' as it is a regulated market and there is a strict limit on the number of registered taxis i.e. there are very strong barriers to entry. The same level of barriers do not exist in road haulage.

Scottish Interurban bus services – Scottish Citylink – anyone else? There are some routes that have a number of competitors, hence some exceptions do exist.

The British Public Transport Market – a relatively high number of companies are involved in this market and a vast number of buyers. At the end of the day a journey is a journey, hence a degree of homogeneity, and if you shop around you should know your prices and that all forms are 'safe' i.e. have had to pass some form of test.

What we have hopefully learned is that the wider you define an industry the more competitive it is likely to be, hence the British public transport market is far more competitive than the Glasgow Bus Market. A second point that we may have learned is that although perfect competition does not exist in reality, this exercise should hopefully have underlined that it does provide a scale (literally in this case!) by which to judge markets and industries with regard to the extent to which they match the 'ideal'. You will have a list that basically reads from left to right those furthest and those nearest the 'ideal'. You should also have a list of reasons as to why you have placed them where you have done. It might not agree with my list or my reasons but hopefully it is not too far away. We could discuss these differences and arrive at some kind of consensus of agreement, which would then give us a common scale. This would then allow us to analyse in a far more constructive and targeted manner the industries concerned, which could then be used in a wide range of purposes – for example a masters dissertation, a consultancy, a highly valued (!) piece of advice etc etc – particularly on how we might move those on the left closer to the right, or alternatively areas that would have to be changed to induce more market driven behaviour from those industries on the left. Let's not get carried away however, and remember that this is a simple introductory exercise, these points are raised simply to illustrate the 'usefulness' of studying something that doesn't exist! (perfect competition).

ANSWERS TO THE UNIT 4 CASE STUDY

Question:

To what extent do you think that the 'vision' that was foreseen through the 1985 Transport Act meets the conditions of perfect competition?

ANSWER

We could be bullish about this question and say that in no manner did the 'envisaged' industry structure meet the market conditions of perfect competition, as perfect competition is an ideal rather than a reality. In fact it has been sometimes compared to the quest for the holy grail, as the search continues but it is never found! Anyway, being more thoughtful we could say that there are several aspects that are similar. Taking assumption one, many buyers and sellers. At the time of privatisation, there were certainly many buyers in the market (100ms), and it was envisaged there would be 'many' sellers – up to around 150 different bus companies. Furthermore, there would be competition present in most areas. Taking Bristol as an example, the former NBC subsidiary would compete with the former local authority company. Secondly, a homogenous product. A bus journey is a bus journey is a bus journey, hence we could say it is a homogenous product. Perfect information – this one is not really applicable, although one could perhaps argue that the vision was that a more customer focused industry would keep the passenger better informed – the reality of course was that the exact opposite happened, but we are only concerned with 'the vision' here. Finally, no barriers to entry and exit. At this point it is worth citing a comment made by the Minister of Transport at the time of privatisation, Nicholas Ridley, who stated that all of the miners that were being made redundant at that time could use their redundancy payments to buy a bus and become entrepreneurs in the bus industry. In a nutshell therefore, barriers to entry were relatively very low, to such an extent that even Nicholas Ridley could fathom that one out, even if the rest of his grip on reality may have been questionable!!

At the very least, it should be clear from the answer that the model of perfect competition was the principle that the privatisation of the industry was working from.

Question:

One of the main reasons cited for the growth in bus company size after 1985 was due to economies of scale – what are the main economies of scale in bus operation and do you believe these to be significant?

ANSWER

It response to this question, it may be useful to distinguish between the sources of returns to scale and economies of scale, and relate these to the issues that were raised in Unit 3 of the course notes. Beginning therefore with returns to scale, these were listed as specialisation of labour, scheduling of inputs, capital inputs and indivisibilities. The specialisation of labour, these will be exhausted at a relatively low

output size, once you have a driver, some admin staff, that's about it really?? Scheduling of inputs – again moving from having a single bus to two buses may not necessarily require you to employ twice as many drivers, as the increase in fleet size may allow for better scheduling of the inputs. There are some gains here therefore, but as above these would probably be exhausted at a fairly low level of output. Capital inputs – buses are not particularly expensive, hence little RTS here. As regards specialisation, productivity gains may be achieved as firm size grows as this would allow the firm to have better suited buses on particularly routes – the most obvious division being minibus/double deckers. There are however also midi, SLFs (super low floor) and single deckers. There are potential productivity gains here possibly over a reasonable range of fleet size before these are exhausted – for example, we do not have to put on a full sized double decker bus to a route that has a small demand, we can employ a smaller more fuel-efficient bus. Finally under this heading is spares and maintenance, which again there may be some potential productivity gains for larger firms as they would carry more spares and hence the 'downtime' of the bus fleet would be less, hence more buses are employed in the carriage of passengers. The final source under RTS is indivisibilities, and this is not applicable – a bus is a minimum requirement to run a bus service and that in itself is a pretty large indivisibility.

Turning to Economies of Scale, listed in the course notes is bulk buying, high cost inputs and financial economies. The major source here is concerned with bulk buying, and two particular aspects – fuel and buses. We may expect larger bus companies will receive discounts in the purchase of these two inputs, and these economies may exist over a considerable range. For example, Stagecoach is far more likely to receive a discount on its order than McKindless (who??). That is because the coachbuilder can afford to cut it's price to Stagecoach because what it loses in profit margin it gains in the quantity sold.

As regards the last part of the question, 'we' believe that returns to scale are fairly limited, but economies of scale are fairly significant. Nevertheless, even these 'significant' economies of scale will be exhausted at a level well below the current size of many bus firms, hence cannot be used on their own to explain the large size of some bus companies – Stagecoach, Firstbus and Arriva to be specific! In order to do so, we would also have to examine other issues, and in particularly, market position.

Question:

What do you see as the main barriers to entry to the bus market?

ANSWER:

In the unit, the main barriers to entry were listed as firm size, high sunk costs, legal protection and control of the factors of production. Ignoring the first one for now, there are no high sunk costs – if your firm goes bust, your main assets – the buses – can easily be used elsewhere, hence relatively easily sold. There is some legal

protection on subsidised routes but you wouldn't want to run on them anyway as you would make a loss, and finally there is very limited control of the factors of production – there may in some parts of the country be a shortage of suitably qualified drivers, hence you may have to pay a slightly higher wage, but that's about it. At this point therefore, there would appear to be few barriers to entry. Which leads us to firm size and the previous question, economies of scale. This is the only real barrier to entry, as the firm would have to enter the market at a reasonably large size (let's say with a fleet size of 500 purely for debate), in order to capture most of the economies of scale – how many bus markets however could withstand an extra 500 buses?

There are numerous examples however where smaller firms have competed successfully against larger operators, and Cole (1998) highlights some of the reasons why this is the case. Firm size therefore may not actually be an issue, hence few barriers to entry. All you need is your redundancy money and away you go….

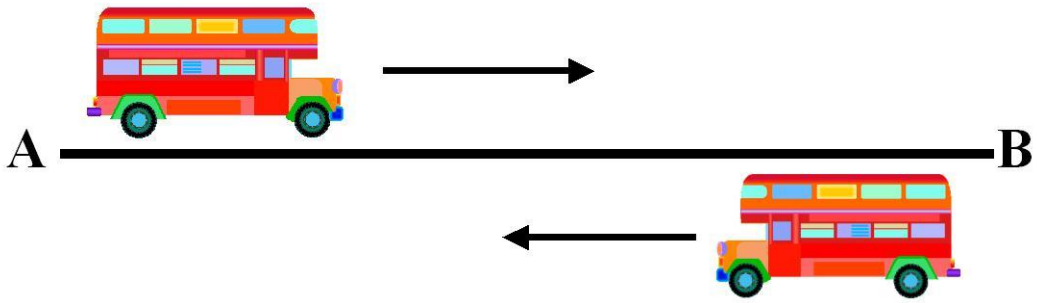
Question:

Do you believe that any of the advantages of monopoly apply in this case study?

ANSWER:

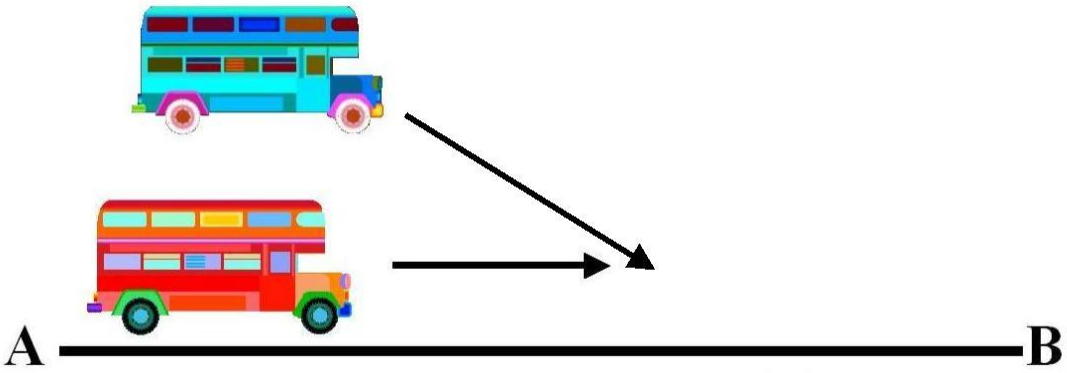
There may be an element of natural monopoly – if you think of the post office, it makes far more sense to have one postal company, rather than have two competing companies. Say for example one company delivers mail to house numbers 1,2,4,5 and 8 and the other to numbers 3, 6 and 7 in a particular street. The first company's delivery person therefore walks past houses 3, 6 and 7, when in fact the marginal cost of delivering to these houses would be very small (as you are walking past the door anyway). There may be some aspect of this with bus services, as small detours would only incur relatively low costs. Of the other advantages listed in the unit, these don't really apply. There is however one further advantage, which again has aspects of a natural monopoly.

Consider the following case, where we have one company operating services between town A and town B:



The company puts on two buses, the distance between A and B is 10 miles and the average bus speed is 20mph. If one company runs this service, then it would make sense to send one bus out from A and one bus out from B, and then they run back and forth all day. Bus frequency would therefore be one bus every half hour.

Now, let's change it, and say that the two buses are run by different bus companies. Bus service patterns are more likely to be like:



The red bus company, sometimes referred to as the 'Black Wheel Bus Company', sets it pattern to run between A and B. It uses its single bus and the frequency for route users would be one bus an hour. It may 'expect' that the green bus company, similarly sometimes referred to as 'the White Wheel Bus Company', to provide the corresponding 30 minutes service. In this case however, it makes more sense for White Wheel to run their bus just in front of Black Wheel's bus, thereby collecting all the passengers along the route and leaving nothing for Black Wheel. Black Wheel will of course respond by re-scheduling their service to run just in front of White Wheel, and so on and so on. This is known as 'Hotelling's Law' first identified by Harold Hotelling in 1929, and occurs more generally where competitors that have 'moveable' vendor outlets tend towards a central point, such as is the case with ice cream sellers. The net result in this example is constant timetable changes and only

one bus per hour for route users. It would make more sense therefore if this route was operated by a single company.

Question:

What does this case study tell us about government intervention in the structure of transport industries?

ANSWER:

In simple terms that it has not been particularly successful, as it has failed miserably to deliver the 'vision' that was originally conceived. This started very early in the process when the MMC had difficulty in defining exactly what 25% of the 'market' meant – the local market?, the national market?, the European market? As stated in the text, In something like 33 referrals to the MMC, very few resulted in any concessions and in most cases the merger was allowed to go through unabridged. That is why today we have three very large bus companies that dominate most of the market.

ANSWERS TO UNIT 5 SELF ASSESSMENT EXERCISES

Exercise 5.1

Briefly summarise the major similarities and differences between an oligopoly and monopoly market structures. Then illustrate, without reference to the notes, the profit maximising position for an oligopolistic market.

ANSWER

Probably the best way to approach this question is to look at the main assumptions of perfect competition and apply these assumptions to these two industry structures to highlight any differences or similarities.

Assumption	Oligopoly	Common	Monopoly
Number of Sellers	Few sellers, many buyers		One seller, many buyers
Product	Differentiated or branded		One product
Information		Imperfect	
Barriers to entry		Significant	

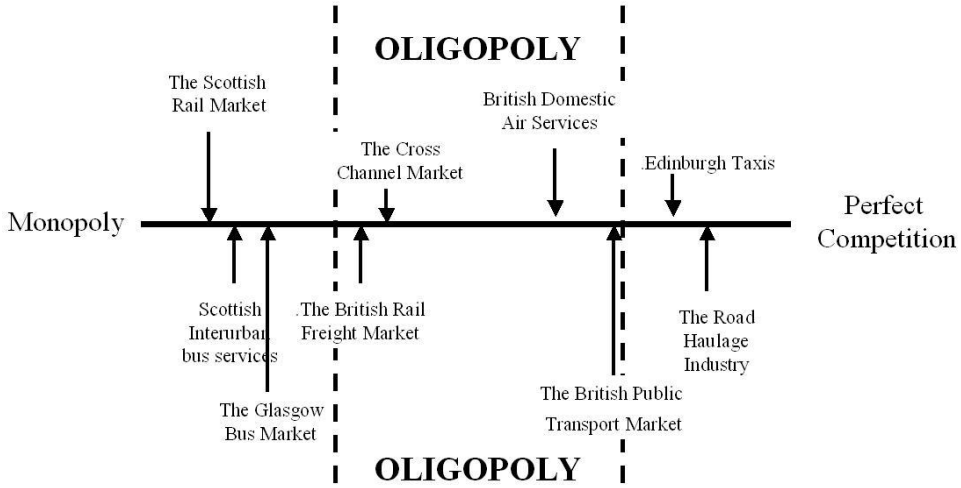
In many ways the two industry structures are very similar, however there are some important differences. Note that under oligopoly it can be that the product is absolutely identical – petrol is a good example of this – however in 'our' model of oligopoly this would be differentiated between through branding. Note that the first two issues (in which they differ) nevertheless is what gives the monopoly/oligopoly firm it's strong position in the market, whilst the second two are what protect them in the long run from future competition (although information will have some bearing on current market position).

Exercise 5.4

Refer back to self-assessment exercise 4.2. On your 'degree of competition' scale, where would you place oligopoly? Of the nine industries/markets listed in that exercise, which three do you believe most closely resemble an oligopolistic market structure? Why have you chosen these three?

ANSWER:

Your diagram should look something like:



The one thing I forgot to mention was that you could re-evaluate some of your earlier positionings from the original exercise! Oligopoly as such would cover a range rather than a specific point. What is not portrayed by the diagram however (due to the nature of most transport industries), is that oligopoly is actually far nearer monopoly than it is to perfect competition. Rather than the middle ground therefore it should be far nearer the left hand side of the diagram. As regards the three most closely resembling such a market structure, these would probably be the cross channel market (as in the full cross channel market), British Domestic Air Services and finally the British Public Transport Market. Note that to the left of our upper boundary we are getting into near monopoly territory, and certainly one in which there are only one or two major operators. The British Rail Freight industry for example has only four companies of note working in the industry, and of these four, two account for most of the market. In our definition therefore, this would constitute an oligopoly, i.e. few sellers, although technically it is a special case of oligopoly that would be known as a duopoly - a market in which there are only two sellers. Probably the best-known duopoly in Britain is the washing powder and

detergents market - despite all those brands there are only two companies, Procter & Gamble and Unilever.

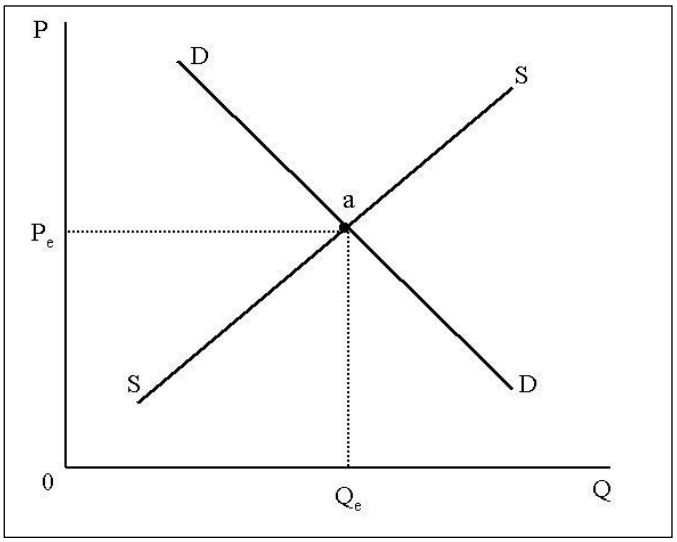
The reason for choosing these three therefore, is that they most closely resemble the overall conditions that were outlined in the unit for oligopoly – few sellers many buyers, branded or differentiated product.

ANSWERS TO UNIT 7 SELF ASSESSMENT EXERCISES

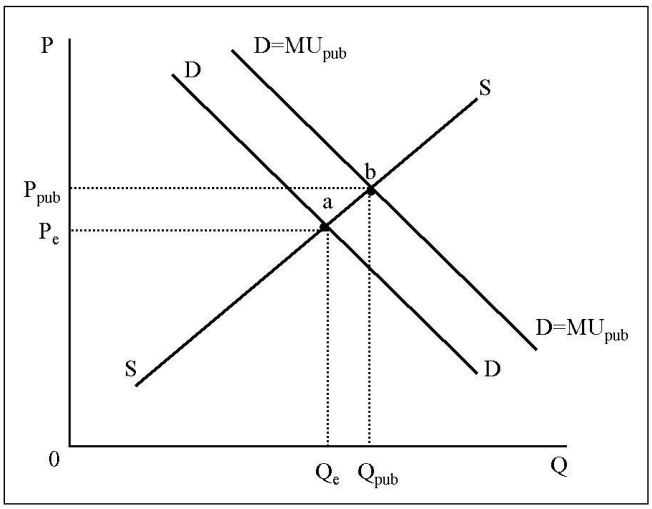
Exercise 7.1

OK, let's have a look at this by firstly specifying the reasons behind paying subsidy to public transport. Most of you will have already completed the Public Transport module (be71012). In that module you will have learned that public transport has three important roles in the transport sector. Firstly, it provides transport for those that do not have access to private transport, secondly it is an efficient method of moving people and finally there are certain circumstances under which people would rather not drive e.g. social factors, long commutes etc.

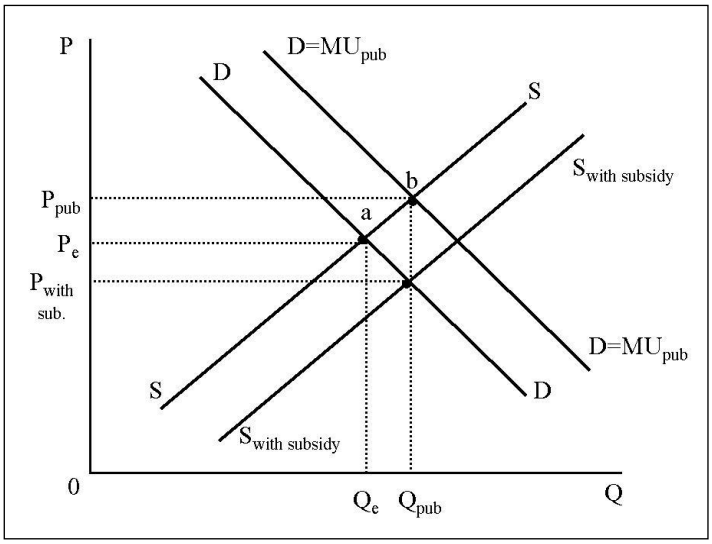
Examining these three roles it is the second one that is of relevance here. By an 'efficient' method of moving people, this breaks down into four further components – land use issues (congestion), pollution, safety and finally it underpins the economy in terms of the labour market. All of these factors are not registered in the market place i.e. there are not direct financial benefit/costs associated with these to the consumers and producers involved. The benefits/costs fall on others i.e. externalities. If we take the basic market position for 'public transport', then this is shown in our basic market diagram:



This is our basic diagram. We know however that not all the costs and benefits associated with public transport are reflected in the demand and supply costs. With public transport, this mainly relates to the demand side of the equation. If more individuals use public transport, then others, i.e. existing road users, will also benefit. These benefits therefore should be reflected in the demand curve. This is shown below:



What this shows is the level of under-consumption of public transport – if all factors were taken into account, then the level consumed should be Q_{pub} rather than Q_e. In order to correct this market failure therefore, the authorities have two options. They can either manipulate the demand side of the market or the supply side of the market. In most cases it is far easier to manipulate the supply side of the market by paying operators a subsidy. This is shown below:



The effect of the payment of subsidy is to move the supply curve to the right as this has the same effect as a decrease in costs. The subsidy is paid to operators to correct for the market failure of external benefits that are not being registered in the market place (under consumption). The net outcome is that the price is reduced from Pe to $P_{with\ sub.}$ and the

new market equilibrium is at Q_{pub}, which is the market 'efficient' position. The subsidy payment therefore has exactly offset the market failure.

Exercise 6.2

The transport authority wishes to implement a concessionary fare scheme. The current demand for bus services from those that would benefit from the scheme is estimated to be in the order 3,500 passengers daily and the flat fare charged by the operator is £1.20. The authority is considering the following options.

Scheme 1: The concessionary fare to be implemented would be a flat 25p charge. This would be expected to increase demand of those entitled to the concession by 10%

Scheme 2: An entirely free concessionary fare scheme which would be expected to increase demand of the affect group by 25%.

a. Calculate for both schemes the expected level of compensation that will be due to the operator.
b. The operator argues that the scheme will prove very popular and if not met by an increase in frequency that this will result in overcrowding of bus services. They argue that the level of compensation being proposed is insufficient to provide for either commercial losses through full fare passengers being left at the bus stop due to overcrowding or the added cost of the extra services. What do you do?

ANSWER

a. Reimbursement

Scheme	**Scheme 1**		**Scheme 2**	
Loss from existing passengers	3,500 * £1.20	£4,200	3,500 * £1.20	£4,200
Cost of carrying generated traffic	3,500 * 10% * 0.20	£70	3,500 * 25% * £0.20	£175
Weekly Totals		£4,270		£4,375
Annual Cost		£1.24m		£1.60m

b. The obvious answer is Scheme 1 as that is the only one within its budget. The authority however may want to consider implementing a free scheme, as given it is proposing on spending such a large amount in the first place the question would be is the 'extra' amount required worth introducing a free scheme?

Panic and complain bitterly that you are not being paid enough to take such decisions! More seriously, there are a range of options but clearly this surrounds the level of compensation to be paid for the extra costs in carrying the generated traffic. What you should probably do is negotiate from the position that ultimately a condition of the operator's license is that they carry concessionary passengers.

As seen however, that might not work! To confirm the operator's figures, you should plug the numbers into the equation and then calculate the percentage increase in order to get the generated traffic. In part II, you should take a worse case scenario and use the operator's generation factor figures to devise some form of scheme and compensation measures that will come within the authority's budget. A 50p scheme with a 35p marginal cost for each extra passenger will cost £1.17m, thus just come in under the budget. In terms of what you should have learned from this last part of the exercise, the key issue is that the economics of the whole operation need to add up, and any transport authority need to understand the underlying economics before implementing any concessionary scheme.

Exercise 6.3

The main question here is the first, but even that we won't go into in any great detail as we have actually gone a long way to answering this first question already. The basic premise of the statement is that high subsidy is normally associated with 'inefficient' operations, hence high subsidy is a 'bad' thing and low subsidy is a 'good' thing. In Exercise 6.1 however, we showed the economic rationale behind the payment of subsidy to public transport operators. This would strongly suggest that the amount of subsidy that has to be paid is simply the amount required to correct for the market failure, irrespective if this is 'high' or 'low'. In simple terms therefore, the statement is false. It should also be noted that we would also have to consider what the political goals of the state were – it may decide, as a political decision, to heavily subsidy public transport for whatever reason it sees fit (e.g. to combat congestion, pollution, social inclusion etc etc etc). The point being that this is a political decision, not an economic one.

The statement does nevertheless highlight another important issue, and that is with regard to the value for (public) money. Economic research would suggest that companies that are heavily subsidised tend to be less efficient. With regard to transport however, and in particular public transport, there is however a causation problem. If we take the case of the British train operating companies for example, those companies that are heavily subsidised tend to be those that operate a high percentage of rural lines. These tend to be lightly used lines, i.e. under utilised, and hence represent inefficient operations. Rather than high subsidy causing inefficient operations, it is inefficient

operations that lead to high subsidy. The important point being that this is a political decision to keep these lines open.

Setting that aside and returning to the value for money issue, this is why many countries have moved to a more competitive industry market structure with regard to the provision of public transport. Most have not gone as far as Britain re a totally de-regulated market, but many have introduced competition for the market i.e. contestability. Again economic research has shown that the main issue with regard to efficiency is not necessarily the transfer of ownership from the public to the private sector, but that the introduction of competition tends to have a far higher impact.

Rather than totally disregard the original statement out of hand therefore, we can perhaps conclude that high subsidy is not a 'bad' thing as long as value for money is ensured.

The other 'quotes':

These should be all simple and fairly straightforward and have been put in simply just to get you to think about some issues, hence:

"Transport should only be subsidised on the basis of what can be afforded by the relevant authorities".

Absolutely true. The only exception could be where the enhancement of transport provision may be seen as an investment, and hence economic growth may be expected to follow i.e. as a supply led measure. Even that however could be argued to still be consist with the original statement, as the authority would (over) spend now in the expectation of future revenue flows in the form of increased tax take.

"'Public' Transport is a 'public' good therefore should be subsidised."

Absolute rubbish, public transport is not a public good hence this is not a rationale for subsidising it. Other rationales for the subsidisation of public transport exist.

"Road networks, because they are provided and maintained by the state, are therefore not subsidised"

Again, absolute rubbish. There is not, and there should not, be a relationship between the revenue raised through fuel duty and road tax and the level of expenditure on the

roads. That would be a normal economic good and one of the main reasons why roads are provided by the state is that they are not 'normal' economic goods. Provisions for road transport should be provided on a need basis. Roads generally do not 'pay their own way' and hence cost the government more than is raised through the tax system, hence are subsidised. At a local level, the situation is even clearer. Urban roads are maintained and provided by the local council, hence are completely subsidised. The issue of congestion charging however is a different issue, as is the issue of users paying the full cost (i.e. including the impact on the environment) of their transport needs, particularly use of the private car. Both of these topics however are for a different discussion.