TRANSPORT ECONOMICS

Only study guide for
TRL2602

Department of Transport Economics, Logistics & Tourism
University of South Africa
Pretoria

Printed and published by the
University of South Africa
Muckleneuk, Pretoria

TRL2602/1/2012–2018

98829440

InDesign

PRTour-style

CONTENTS

OVERVIEW

1. A WORD OF WELCOME

We have great pleasure in welcoming you as learner to the module on TRANSPORT ECONOMICS and trust that you will find your studies in this module interesting, stimulating and instructive.

In this module we want to look at the different links between transport and the micro-economic theory. You will see that Transport Economics is a large field of study. However, we are going to look at the critical aspects of Transport Economics and how these can be applied in reality. Once again welcome to exploring the world of transport.

2. THE PURPOSE OF THIS MODULE

To enable learners to comprehend the essential economic aspects of transport so that they can contribute towards the implementation of economic efficiency in transport.

The purpose of this module is to introduce you to Transport Economics theory and the theoretical concepts that are applied to transport. The economic and commercial aspects of transport are largely determined by its unique technical and operating characteristics. The meaningful and successful application of economic and management theory requires an in-depth knowledge of these technical and operating characteristics of transport. It is impossible, within the scope of a single module, to familiarise you with all the important technical and operating characteristics of transport. Hence we have worked selectively to include in this study guide those characteristics that are particularly important to the course in transport economics as we present it.

Learning assumed to be in place and recognition of prior learning

The credit calculation is based on the assumption that students are already competent in terms of the following:

- **Literacy:** Acceptable levels of reading and writing proficiency as required by first level tertiary education in the main language of tuition; as well as fundamental numerical skills, including basic arithmetic.
- **Knowledge:** An informed understanding of one or more fields within the transport industry and an awareness how the industry relates to the business environment.
- **Gathering information:** An ability to gather and analyse information by effectively selecting and consulting relevant written and electronic sources.
- **Communication:** An ability to present and communicate information coherently and reliably in written, oral and electronic formats, while using academic and professional discourse, in the main language of tuition.
- **Independence:** A capacity to take responsibility and evaluate personal learning needs; and take the initiative to address strengths and weaknesses thereof, within a structured learning environment.
- **Technology:** A capacity to effectively and efficiently access and utilise essential electronic resources, specifically relating to the use of computers and the Internet.

Range statement for the whole module

The following scope and context apply to the whole unit standard which links transport and the transport theory and upon completion students will have a clear understanding of the critical aspects of transport theory and how these can be applied in practice.

Outcomes, assessment criteria and range statements

A range of tasks in the study guide, tutorial letters, assignments and examination will show that students have achieved the following outcomes:

Outcome 1:

Demonstrate an understanding of the theoretical concepts of transport economics.

Assessment criteria:

- Define relevant concepts such as transport economics, empirical object and study object of the discipline, utility
- Explain how transport economics relate to economics
- Explain the function and nature of transport
- Analyse the economic developmental role of transport
- Identify the factors where transport has a role to play, like social, political and strategic role

Outcome 2:

Explain the workings of the transport market in micro-economic terms, indicating the differences between transport demand and supply vis-à-vis demand and supply of manufactured products.

Assessment criteria:

- Explain why a need for transport exists as well as related aspects such as utility and demand elasticity
- Distinguish between different needs to transport goods versus needs to transport passengers
- Indicate the differences between the demand for transport and the demand for manufactured products
- Explain the role and importance of the derived nature of transport demand
- Describe the principal characteristics of transport demand
- Explain why transport supply should be viewed differently from the supply of manufactured products
- Indicate the differences between the supply of transport and the supply of manufactured products
- Explain the supply function of transport, perceived cost and the components of this function
- Explain the factors determining the characteristics of transport supply

Outcome 3:

Demonstrate an understanding of the integration between transport demand, supply and costs to form the economic structure of transport.

Assessment criteria:

- Define various costs relevant to transport, like total, average, marginal, fixed and variable, joint and common, direct and indirect costs
- Indicate how economic costs differ in nature from financial costs
- Explain the behaviour of costs in the cost structure of enterprises and how economies of scale, economies of density and economies of scope come about
- Distinguish the cost behaviour of a multiproduct enterprise like transport, from an enterprise producing a single homogeneous product
- Explain the role user costs play in the transport markets
- Analyse the role played by coordination, economies of scale and scope in determining the size of transport enterprises
- Identify the attributes of the transport market
- Briefly explain the various forms of competition found in micro-economics
- Explain the concepts of workable competition, contestable markets and ruinous competition and how these relate to the transport market
- Indicate the forms of competition found within the different modes of transport.

Outcome 4:

Demonstrate an understanding of external costs of transport and pricing in transport

Assessment criteria:

- Explain what is meant by an externality and external costs in economics
- Briefly describe the nature of external costs
- Identify typical externalities transport causes towards its environment like pollution and congestion
- Explain how the external costs of noise pollution should be integrated into the total costs of transport by means of the optimal pollution charge
- Explain what is meant by transport congestion as well as the negative consequences it has for the users of transport
- Indicate, both graphically as well as algebraically, how the congestion charge can be used to represent the true economic costs of the excessive use of a limited transport facility
- Explain what is meant by economic efficiency
- Identify the reasons why pricing according to the marginal cost of the output would lead to economic efficiency

- Indicate graphically and by means of written explanation how second best pricing will enhance social welfare if marginal cost pricing is not followed
- Explain the requirements, types of and application to transport enterprises of the economic principle of price discrimination
- Explain graphically the problems of pricing in transport when there is a high proportion of joint costs

Outcome 5:

Demonstrate a basic understanding of how investment in transport infrastructure can influence social welfare of the community

Assessment criteria:

- Explain the differences between the private versus the public (social welfare) approaches to investment graphically
- Indicate the different approach and the attendant difficulties in investing in indivisible infrastructure
- Explain cost-benefit analysis as a tool to measure the economic suitability of a transport investment
- Graphically indicate the benefits derived from a transport investment and indicate the types of benefits and costs associated with such an investment
- Explain the national income approach as an alternative to cost-benefit analysis

The module also helps learners to achieve the following Critical Cross field Outcomes as formulated by the South African Qualifications Authority (SAQA):

The ability to:

- Identify and solve problems in a way that demonstrates responsible decision-making ability, using critical and creative thinking
- Work effectively with others as a member of a team, group, organisation or community
- Collect, analyse, organise and critically evaluate information
- Demonstrate an understanding of the world as a set of related systems by recognising that problem-solving contexts do not exist in isolation
- Develop entrepreneurial opportunities

3. THE PLACE OF THIS MODULE IN UNDERGRADUATE STUDY

Undergraduate studies in the College for Economic and Management Sciences at UNISA provide graduates with a wide range of knowledge and skills that will equip them to work in the business community.

This module is compulsory for the following degree

The BCom degree with specialisation in Transport Economics

Apart from being a compulsory module for the above degree, TRL2602 is also compulsory for a major in Transport Economics or Logistics as part of a general BCom or BAdmin degree. The module can also be chosen as an elective in other BCom or BAdmin degrees.

This module provides a foundation for further studies in Transport Economics and is indeed a prerequisite for most third-year modules in Transport Economics or Logistics. Students who have completed the module will probably continue their studies in Transport Management, Transport Planning and Investment, Air Transport, Sea Transport, Rail Transport and Road Transport majoring in Transport Economics or Logistics also makes it possible for students to engage in postgraduate studies (ie an honours degree, master's degree and ultimately a doctoral degree).

4. STRUCTURE OF THE STUDY GUIDE

A systems approach

As you work through this study guide, you will realise that the transport process interacts with its environment to such an extent, and that the activities and resources that make up the transport process are so closely related, that they in fact form a system. A system can be described as a collection of elements which are related or interdependent and which consequently form a complex unit. It is therefore sensible to adopt a systems approach to transport in general and to transport management in particular. This ensures that we identify all the factors or variables that affect the coordination of the transport process, establish the connection between these variables and determine the effect which coordination or management decisions have on the system as a whole rather than just a part of the system (or subsystem). Decisions which take only a part of the system into account usually benefit that subsystem only and result in sub-optimisation, whereas decisions should be to the advantage of the entire system.

Approach to the teaching of transport economics

In this module we use a combination of approaches to the teaching of Transport Economics.

Firstly, we use a systems approach. This means that we consider the environment in which transport takes place and the transport process itself as subsystems of the transport system. We indicate the interaction between the different subsystems and between the components of the subsystems.

Secondly, we discuss transport theory with reference to the main functions of a transport enterprise. We concentrate on the transport processes, technology in transport, network design and planning.

Thirdly, we follow a consultant approach. In other words, we analyse the transport systems and modelling.

Outline of the study guide

In accordance with the above approach, and to achieve the purpose of the module, we have structured the study guide as follows:

This study guide consists of three main parts which are subdivided into study units, of which there are a total of 10. The first main part consists of three study units. Part 1 is headed, Introduction to Transport Economics. In this part, Transport Economics is placed in perspective, and you will become acquainted with Transport Economics and why it is studied. In study unit 1 we provide you with an overview of how Transport Economics developed. In study unit 2 the relationship to Economics is indicated and we stress the importance of Economics as a source of knowledge for Transport Economics. One of the principal characteristics of transport is that it is a means to an end, and not an end in itself. Thus transport is studied because it can lead to development. In study unit 3 the economic developmental role of transport is briefly explained.

Part 2 of this study guide deals with the Transport market. It is extremely important that, when studying Transport Economics, one understands the functioning of the transport market. The two most important elements of a transport market are demand and supply. It is in fact the interaction between demand and supply in the market which is of cardinal importance for the efficient operation of the transport market. Consequently, the demand for transport is discussed in study unit 4. In study unit 5 the "other side of the coin" is considered, namely the supply of transport. For the purpose of determining transport supply, the cost of providing the transport service is of cardinal importance. This is why transport costs are discussed in study unit 6 and why the most important cost concepts are explained. The interaction of costs gives rise to the cost structure of an enterprise, and the cost structure together with the demand for transport services in the market gives rise to the market structure of transport. This is why the economic structure of transport is explained in study unit 7. Within this structure, both supply and demand factors play a role in the determining of prices or tariffs in the transport industry. In study unit 8 we deal with something which is also important for the determination of transport prices, namely the external costs of transport. Transport gives rise to externalities such as pollution and congestion, and these externalities must be taken into account when determining prices. In study unit 9 we discuss price determination in transport, that is, the determining of prices for both freight and passenger transport.

Part 3 consists of only one study unit. In the sphere of transport we distinguish between moving units, such as vehicles, vessels and aircraft, and the infrastructure, such as roads, railway lines, stations, harbours and airports. In study unit 10 we discuss infrastructure investment as well as a technique for evaluating investment, namely cost-benefit analysis.

5. RESOURCES YOU WILL NEED

The prescribed material for this module consists of:

This study guide

Tutorial 101 (which you should have received together with the study guide when you registered)

Tutorial letters containing answers or suggested solutions to assignments

Any other tutorial letters that may be sent to you during the course of the semester

There is no prescribed book for this module. The study guide is your main source.

You can also visit the following websites: *www.transport.gov.za, www.spoornet.co.za, www.transnet.co.za, www.google/transport.com www.saa.co.za* and *www.saponet.co.za*, and get an update of what is happening in the field of transport. Some of the daily and weekly newspapers may also provide valuable information.

6. GUIDELINES FOR STUDYING THIS MODULE

Time frame for studies

As indicated above, the module consists of 10 study units. Try from the start to master a study unit per week. This should leave you with sufficient time to revise and prepare for the examination.

Learning outcomes

Each study unit commences with a purpose statement and learning outcomes. These outcomes are based on the outcomes for the module, as formulated at the start of this preview, and give you an indication of what you should be able to do once you have studied the particular study unit.

Activities

Within each study unit there are a number of activities which you are instructed to do. The activities will help you to master the study material and to apply it in practice, thereby enabling you to achieve the outcomes of the study unit. Many of the questions asked in the activities are similar to the questions you can expect in the examination. It is therefore to your own advantage to do the activities.

Self-evaluation questions

Each study unit ends with self-evaluation questions which you must be able to answer. These are aimed at finding out whether you have achieved the purpose and outcomes of the study unit, and they are of particular importance in preparing for the examination. Also refer to the section on assessment of the module (see below). Develop an understanding of the concepts and ideas.

To achieve the purpose and outcomes of this module, you need to develop a deep understanding of important concepts and to learn how to apply these concepts in real-life situations. To develop your understanding, you must deliberately process the information you are reading. As you read, try to identify the main ideas, think about why they are important, think about how these important ideas are related to one another, and think about the implications they have for people in the real work environment. You can also test your understanding by trying to summarise the ideas, trying to explain them to other people and trying to find examples of them operating in real-life situations.

7. ASSESSMENT OF THE MODULE CONTENT

There are three types of assessment for this module: Self-evaluation questions in the study guide; self-evaluation assignments or compulsory assignments included in Tutorial Letter 101; and an examination.

Self-evaluation questions

As mentioned above, each study unit concludes with self-evaluation questions. These questions are typical of the sort of questions you can expect in the examination. If you write down your answer to each, you will be able to check your answers and reasoning by referring back to the relevant section in the study guide. This self-assessment is one of the most effective ways of learning the things we want you to learn in this module.

Self-evaluation assignments and compulsory assignments

Self-evaluation assignments are provided in Tutorial Letter 101. These assignments are designed to help you learn and to provide you with feedback. Guidelines for assessing the self-evaluation assignments are contained as an annexure to Tutorial Letter 101.

If there are compulsory assignments in Tutorial Letter 101, you must submit them to the Assignments Section (in the envelope provided) before the due date. They will then be marked by your lecturers and returned to you with comments. More detailed instructions pertaining to compulsory assignments will be given in Tutorial Letter 101.

The examination

The examination is your opportunity to show that you have achieved the outcomes of this module. A specimen examination paper is provided in Tutorial Letter 101 to give you an indication of how the paper is structured.

Key terms in assessment

When you assess yourself, do compulsory assignments or answer examination questions, take note of the following meanings of assessment terms:

List:	Write single words
Define:	State the meaning and mention the essential qualities
Discuss:	Identify the issues and provide points for and/or against particular issues
Describe:	Provide details of characteristics and features

Explain:	Make clear; describe cause and effect; make the relationship between things evident; provide details of why and/or how something occurs
Compare:	Show how things are similar
Contrast:	Show how things are different
Analyse:	Identify components and the relationship between them; draw out and relate implications
Evaluate:	Make a judgement based on criteria; determine the value of something
Substantiate:	Give reasons

8. ACKNOWLEDGEMENTS

We acknowledge contributions by the following person to previous study guides for this module and for work done in developing the curriculum of the module over a number of years:

The late Prof CA Smuts

TRANSPORT ECONOMICS AND LOGISTICS

OVERARCHING AIM

To enable students to identify and discuss the essential economic characteristics of transport.

PART 1

Introduction to Transport Economics

OVERVIEW OF PART 1

The aim in this part is to give you, the student, an idea of where Transport Economics as a discipline fits in. Transport Economics is an applied science that applies the principles of economics and business management to transport. You have already become acquainted with subjects such as Economics and Business Management, but, in this part, we shall endeavour to help you highlight the position of Transport Economics within the economic and management sciences.

Consequently, we shall provide you with a brief historical overview of the development of Transport Economics, and will indicate the close relationship which exists between the disciplines of Transport Economics and Economics. You will discover that micro-economics in particular is a source of knowledge for Transport Economics. Thus a sound basic knowledge of micro-economics is a prerequisite if you wish to be truly successful in your study of the subject Transport Economics.

STUDY UNIT 1

The development and field of study of Transport Economics

CONTENTS

OBJECTIVES

After having studied this study unit, you should be able to:

- explain how the concept "transport economics" originated
- provide a brief overview of the development of Transport Economics as a field of study
- indicate the relationship between transport economics and other economic sciences
- explain why we study Transport Economics and where it finds its practical application

1.1 INTRODUCTION

Why is transport necessary? Could society not in fact get by without transport? Why must transport occur economically?

Reflect on the above questions for a while. Did you stay in one place today or did you go somewhere? If you went somewhere, why did you go there? Did you walk there or did you make use of a means of transport in order to get to your destination? Did you pay to travel to your destination? If your answer to this last question was "no", are you, subconsciously, perhaps NOT telling the truth?

Transport Economics does not exist in a vacuum, but affects all of us – not always because we want it to, but, in most cases, because it cannot but do otherwise. Let us assume that you travelled somewhere today, even if it was only to take your children to school. Your children had to get to school and you took them there in the family car. Your reason for taking them was to get them to school and not to drive around. You may perhaps think that you paid nothing for the journey – apart from using a little petrol – but your assumption would probably be incorrect. Was your journey an economic one? This will depend on the total cost of such journey. The little petrol which you thought of is your perceived cost but is not necessarily the total cost of the journey.

It is these things, among others, with which transport economists concern themselves. We hope that you will find your studies for this paper in Transport Economics both interesting and of value, since transport occupies a very important place in our daily lives.

1.2 HISTORICAL DEVELOPMENT OF TRANSPORT ECONOMICS

1.2.1 General

The dynamic nature of transport economics in the modern world may give rise to the impression that Transport Economics is a comparatively new field of study. However, this is in fact not the case – for a considerable time it has been an important field of

interest and investigation among persons such as economists, engineers, management scientists, lawyers, geographers and those involved in the provision and use of transport services. The first specialised publication on transport economics appeared in the forties of the previous century, that is, about seven decades after the publication in 1776 of Adam Smith's renowned work "The wealth of nations", which may be regarded as the first academic attempt to study the economy in general. To provide a detailed description of the development of transport economics would not be possible here. Transport Economics as a discipline is of fairly recent origin and may be dealt with as part of a number of different disciplines, or, as we view it at Unisa, as a separate discipline falling within the family of economic and management sciences. Moreover, we also use the term "transport economics and logistics". The all-embracing nature of the concept "logistics", and specifically, "transport logistics", gives Transport Economics its right of existence. In earlier years, Transport Economics was regarded as a part of Business Economics (Business Management) or pure Economics. Although Transport Economics is now regarded as a discipline in its own right it has not relinquished its link with Economics and Business Management. We distinguish between ***Transport Economics*** and ***Transport Management.*** The first is studied in this module and the second in the paper called ***Transport Management (course code TRL201X).*** The role of ***Logistics*** and ***Supply Chain Management*** is not forgotten and also forms a strong link with ***Transport Management.*** The ***Tourism*** industry makes use of transport to get tourists to the attractions they visit and one cannot tour without using transport as the means of touring. Therefore the economics of transport plays a big role in Tourism. At Unisa (University of South Africa) this was realised and you are now studying in the Department of Transport Economics, Logistics and Tourism.

1.2.2 Historical development

Transport economics also originated from the economic analyses conducted in the field of micro-economics. As far back as 1913, Taussig published an article in the Quarterly Journal of Economics on railway rates and joint costs (Taussig 1913:692–699). It was therefore economists who conducted cost analyses of transport problems which in fact gave rise to the development of economic theories. Up to and including the twenties of the present century, the normative analysis of transport problems flourished. By "normative analysis" is meant that the analysis had to establish certain norms, such as what the right levels of passenger tariffs should be and what the social benefits should be of regulating transport providers.

Between the twenties and fifties, interest in the analysis of transport problems declined. Instead, economists concentrated on the mathematical analysis of transport problems and on macroeconomics. After the Second World War, the increased use of private motor vehicle transport, and the mobility which resulted there from, gave rise to renewed interest in the economic analysis of transport problems. Although normative aspects were still of importance, an interest in the positive analysis of transport economic problems also developed. Certain questions such as what determines the demand for and supply of transport, what role costs play in the supply of transport, et cetera, became increasingly important. At the same time, mathematical models and analytical techniques were developed in order to facilitate analyses.

In the meantime, however, modern transport technology was developing rapidly, and large-scale urbanisation and industrialisation occurred. Mechanised transport became increasingly interwoven with the daily social and material existence of society. The increasing importance, scope and complexity of transport in turn stimulated interest in transport matters, and in the late fifties a number of important publications on transport economics saw the light of day. To a certain extent, these works built on the foundations laid by earlier works and dealt mainly with the allocation of economic resources in the sphere of transport. The emphasis was placed particularly on transport costs and on criteria for price determination in order to ensure the economically efficient use of resources and the optimal allocation of traffic to the various transport modes. In the ensuing decade, the field of study of Transport Economics grew rapidly in both scope and diversity. Apart from the fact that existing solutions to transport economic problems had to be continually revised in the light of new factors such as deregulation and technological development, problems associated with, among other things, urban transport, physical distribution and the relationship between transport and economic development increasingly occupied the attention of transport economists. In modern society, the availability of a technically and economically efficient transport system is not only a prerequisite for maintaining and expanding an efficient economy, but also has an important influence on people's quality of life. Accordingly, the study of transport from both an economic and a multidisciplinary point of view is still in the process of developing, particularly as far as content and refinement are concerned. Regardless of the degree of refinement which may already have been achieved with regard to transport economic theory, changing technology and institutional circumstances are forever causing areas of existing knowledge to become outdated, and this in turn requires ongoing research and the adaptation of syllabi for Transport Economics.

De Wit and Van Gent (1996:28–29) explain the changes that have taken place over the years in the European situation that increased the terrain of Transport Economics, specifically in the Netherlands. They also maintain that Transport Economics has changed from only the application of micro-economics to transport as something more comprehensive. **Transport Economics** as discipline is not a sector of economics like **health economics** or **agricultural economics. Transport Economics** straddles the whole economy of a country. They also emphasise the fact that transport politics play a big role and that the approach of welfare theory is followed. Here they refer to the fact that congestion, inter modal transport, external costs and infrastructure problems are prevalent

in modern transport. Hensher and Brewer (2001:4) also show the growing integration of the institutional environment, transport economics and organizational management (logistics, supply chain and business management). When studying Transport Economics you should keep these changes in mind and remember that you cannot study transport economics in isolation.

1.3 THE EMPIRICAL OBJECT AND THE STUDY OBJECT OF TRANSPORT ECONOMICS

Before considering the function and nature of transport, we must indicate where Transport Economics fits in and what the scientific basis is for making such a statement; in other words, why do we study Transport Economics?

The Masakhane Project is in full swing in your community. The mayor and all the councillors have given signed undertakings that they will fulfil their obligations and will pay their monthly municipal accounts. However, you see no need to give such a signed undertaking or to make the relevant payments. The reason for adopting such an approach is that you are most unhappy with the situation regarding the streets in the area in which you live – at present there are no tarred roads and the vehicles of all the inhabitants are continually being damaged. "If they are prepared to do something about the situation, I shall also share in the spirit of Masakhane," you tell your neighbour. Your neighbour thereupon makes an appointment with the mayor for you. The mayor receives you cordially, allows you to put your point of view and indicates that he is sympathetic towards your viewpoint. However, he asks you the following questions:

(1) Which streets should be tarred?
(2) Why should these specific streets be tarred?
(3) What economic benefits will the community enjoy if the streets are tarred?
(4) How much will it cost to tar the streets?

How would you answer these questions?

You may be able to answer the first question off the cuff, or may be able to answer it reasonably well as a result of having had prior discussions with members of the community. However, the other questions will be somewhat more difficult to answer. To answer them, knowledge of Transport Economics is necessary. A local authority which cares for its people will have a section which will consider such matters from a transport economic viewpoint.

The study object of Transport Economics is the optimal utilisation of scarce transport resources in order to provide for unlimited transport needs.

The empirical object of Transport Economics is the entire community which experiences transport needs.

When we therefore speak of Transport Economics, we are in fact referring to that science which considers the transport needs of all people and then determines how such needs may be satisfied with the limited means available. Transport Economics thus studies how unlimited transport needs can be satisfied optimally with limited transport resources. The empirical field within which such a study is undertaken is not the individual transport enterprise (as it is in the case of transport management), but the entire community which makes use of transport or is affected thereby.

Transport Economics as a science can, on the basis of scientific techniques such as cost-benefit analyses for example determine what the cost of the tarred roads will be, which streets it will be best to tar from an economic viewpoint, what this will cost and whether it will result in adequate economic benefits for the community.

Now that you know how Transport Economics developed and what you will be studying, we can examine the relationship between transport and economics in more detail.

SELF EVALUATION

(1) Do you regard Transport Economics as a discipline it its own right? Give reasons for your answer. (5)
(2) Which two other disciplines form the basis for Transport Economics? Which sub-discipline of one discipline is the basis of Transport Economics? (3)
(3) Give a brief historical overview of the development of Transport Economics. (10)
(4) What are the empirical object (study field) and the study object (goal) of Transport Economics? (5)

STUDY UNIT 2

The relationship between economics and transport

CONTENTS

2.1 THE FUNCTION AND NATURE OF TRANSPORT

To place the economics of transport in its proper perspective, one must consider transport itself. We shall now examine the nature and function of transport. In doing so, we wish to refer once again to the example given in section 1.2, that is, to the example of you transporting your children to school. Why did you take the children to school by motor vehicle?

People are continually engaged in a large variety of activities, such as living, working, conducting business, consuming goods and services, visiting friends and family members, and enjoying recreational activities. These activities are usually directed at satisfying needs, which may vary from the most basic needs, such as the need for food, shelter, clothing and security, to higher-order needs, such as the need for social interaction, education and training, health care and recreation. Obviously, few of these activities can be engaged in simultaneously with other activities, and, owing to the particular requirements of different activities, different spaces or geographic areas are adapted in order to accommodate them. Factors such as location, topography, and climate, the availability of labour and raw materials, and historical land-use patterns determine to a significant extent where specific activities will be established or performed. Especially in highly developed areas, such as towns and cities, a large variety of adapted spaces or land uses are encountered, for example commercial, industrial and residential areas, as well as hospitals, schools, theatres, libraries and sports complexes. Furthermore, at regional, national and international level we encounter specialisation in specific activities or land uses, such as agriculture, mining, forestry, manufacturing and tourism. Apart from the fact that few of the activities concerned can occur simultaneously with other activities, many of the activities are performed at fixed times, for example formal schooling and work activities (which, in most cases, begin and end at specific times on working days), holidays, agricultural activities (which are confined to specific seasons), and recreational activities (which are normally engaged in over weekends and in the evenings). Ultimately, therefore, widely diverse human activities are separated from one another as far as time and space are concerned. Participation in these activities consequently requires the movement of people so that they are able to be at specific places at certain times. The spatial or geographic movement of people and goods is achieved with the aid of transport. From the purely physical viewpoint, transport embraces the movement of people and goods between pairs of geographic points using technical means, the aim being to ensure that such people and goods are at specific places at specific times. The departure point is usually termed “the origin” and the arrival point “the destination”.

Technical means are indispensable to transport, and, although they are employed in order to achieve the same goal in each case, namely the movement of people and goods, the technology concerned may differ from one case to another. In this regard, different forms of transport are distinguished on the basis of the medium in or on which they are mainly to be encountered. Moreover, a distinction is drawn between the different transport modes occurring within each form, for example rail, road and pipeline modes in the case of land transport. As far as the technical means are concerned, it is also important to distinguish between infrastructure, such as roads, airports and harbours, which constitute the static elements of the transport system, and vehicles, aircraft and vessels, which are in effect the mobile elements of such system. The transport system also includes auxiliary equipment and systems which are essential for the successful operation of the system. An important function of the transport infrastructure is to link different activity areas with one another, thereby providing people and goods with access to these areas. The function of vehicles, aircraft and vessels, and the auxiliary equipment associated therewith, is to provide (together with the infrastructure) people and

goods with mobility. The flow of people and goods between specific pairs of origins and destinations is known as "traffic". From a functional viewpoint, transport is a process comprising all those activities which are necessary to move people (passengers) and goods (freight) from an origin so that they will be at a specific destination at a specific time. Generally, such activities are directly related to the spatial movement of people and goods, and are carried out successively in recurring cycles, for example the loading of freight at the origin, the transportation thereof to a destination, the unloading of the freight at such destination, and the return journey to the origin with the empty vehicle – that is, in the case of road transport in one direction only. Apart from the movement process, which constitutes the basis of transport, transport also includes a number of auxiliary processes, such as the sorting, consolidation, despatch and storage of freight in the case of freight transport. The auxiliary processes associated with passenger transport are not as obvious, but include those activities which are necessary to deal with passengers who report to or arrive at terminals until they once again depart. The capacity of human activities – which include the provision of goods and services as well as consumption – to satisfy the needs of people means that such activities have a specific value or utility for the persons concerned. Since there is little opportunity for direct need satisfaction in modern communities, activities must be arranged, and goods and services must be created by combining and processing different inputs, before they can contribute to the satisfaction of needs and therefore be of value (or possess utility), for example, during the manufacturing process, raw materials must be converted into products which are able to satisfy specific needs. This type of utility is known as utility of form. Transport, in contrast, creates utility of time and place. In the case of passenger transport, people are moved from places where, at specific times, there is little or no supply of opportunities for satisfying specific needs, or where there is little or no need for their presence, to places where there is a greater supply of opportunities for need satisfaction or where there is a greater need for the particular person's presence. This dual explanation for the existence of passenger transport can be explained on the basis of the journeys undertaken by employees from residential areas to their places of work. Employees who need to earn an income travel at the appointed times from residential areas where, at that particular time, there is a shortage of job opportunities (and consequently of opportunities to earn an income), but an oversupply of employees, to industrial, commercial or office areas where, at that particular time, there is an oversupply of job opportunities and a shortage (and thus a greater need) of employees. For both employers and employees, the fact that the latter can be at their places of work at the specific time required is therefore of value. Similarly, in the case of freight transport, goods are moved from places where the supply thereof exceeds the need for these goods, and where the surplus of such goods is of little value at that particular time, to places where the need for the goods exceeds the supply thereof at that particular time and where the goods will consequently be of greater value. Raw materials are, for example, transported from mines and farms, where there is an oversupply, to industrial areas where they are required for the manufacture of consumer goods. The latter, in turn, are transported to commercial areas where they are made available to consumers who need them. Note that the creation of place and time utility through the transportation of people should not affect such people's intrinsic capacity to enjoy opportunities, or even to be of use themselves. The transportation of goods should also not change their inherent utility (eg the form utility of manufactured goods). In fact, in the case of people, it is most important that the transport process does not result in excessive fatigue, in injury or in death, or in the case of goods, in damage, deterioration or loss. By now you have probably realised that transport is virtually never an end in itself, but is purely a means to an end. This is known as the derived nature of the transport need.

What the last-mentioned term means is that the need for transport is derived from those needs which require transport in order that they may be satisfied. In this regard, journeys are usually assigned a specific objective and we refer inter alia to work, school, recreation, business and social journeys in the case of passenger transport, and to collection and delivery journeys in the case of freight transport. A further consequence of the derived nature of the transport need is that transport users make specific demands of a transport service which accord with the objective for which it is employed. These requirements are referred to as quality or user requirements and have a bearing on, for example, the frequency, speed, regularity, reliability, suitability, safety and convenience of a transport service. Up to now we have concentrated only on that which can be achieved with the aid of transport, that is, the creation of place and time utility. However, the movement of people and goods does not occur instantaneously or without any sacrifices. Such movement takes time and requires the utilisation of scarce resources which can also be used for other purposes. As far as time is concerned, the various activities forming part of the transport process provide virtually no opportunity for the simultaneous performance of actions, and the time taken up by individual activities must be added together in order to determine the total time of a transport cycle. This means that transport takes up people's time, while goods are not available for processing, sale or consumption during the transportation thereof. Furthermore, protracted journeys may result in people being inconvenienced and becoming fatigued, and may cause goods to be spoiled. There are therefore sound reasons why transport time should be kept to a minimum. However, time is but one of the inputs in the transport process. Scarce natural, manufactured and human resources are also utilised in the process. As far as the transportation of natural resources is concerned, large areas of land are required for the erection of terminal facilities such as stations, airports and harbours, or for the construction of roads, pipelines and railway lines. Crude oil and coal, which are also natural resources, provide directly and indirectly for virtually all the energy needs of transport. Manufactured resources – also referred to as capital – are in effect the same as the technical means referred to earlier, namely terminal facilities, roads, railway lines, pipelines, vehicles, aircraft and vessels, as well as any auxiliary equipment used in the transport process. Human resources – or labour – which is used as inputs in the transport process, vary from unskilled labour to people with different types and levels of training and expertise. The consumption of scarce resources represents the cost of transport. Natural resources, capital and labour (together with entrepreneurship) are also generally referred to as the factors of

production, because they are used for the production of useful goods and services. In this study guide, "resources" and "factors of production" are used interchangeably. Viewed narrowly, the physical distance between two geographical points serves as the measure of the spatial divide between them. However, on the basis of the broader perspective discussed in the preceding paragraph, the divide between origin and destination which must be bridged by means of transport may be considered from different viewpoints. Apart from pure physical distance, the divide may also be measured on the basis of the time or cost required to bridge it. For a given distance, a trade-off between transport time and transport cost is in fact possible. Shorter transport times usually go hand in hand with higher transport costs, and longer transport times with lower transport costs, for example it is commonly known that air transport is quicker, but more expensive, than road transport over distances where these two types of transport are able to compete with each other.

To summarise, we may state that the function of transport is to contribute to the satisfaction of human needs by creating place and time utility. The creation of place and time utility is achieved through the movement of people and goods from places where they can contribute relatively little to the satisfaction of needs, to places where they can make a relatively greater contribution. The transport process, which embraces the spatial movement of people and goods, takes up time and requires the use of the scarce resources of the community, while scarce resources, or factors of production, are employed in the process. One complication which arises is that it is not only the transport enterprise's costs, as reflected in its prices, which represents the cost of transport for the transport user. Commuters take into account not only the monetary cost of their journeys, but also the amount of time which such journeys take and, frequently, the quality of the service which is offered. This explains why, for example, many commuters prefer the more expensive, but faster kombi-type taxis to the cheaper, but slower and more inconvenient, buses. Similarly, freight despatchers or consignors are interested not only in monetary cost, but also in the speed, reliability and frequency of freight transport services. Transport is not unique in this respect, but does differ from other products and services in that monetary cost may comprise only a comparatively small part of user cost. In order to understand the economic behaviour of transport users, it is thus important to take into account the full opportunity cost which makes provision for all these divergent cost factors. Usually, the full user cost is referred to as generalised cost. The monetary cost (price) of a transport service, and the monetary value of travel or transport time, normally constitute the most important components of generalised cost. Apart from the fact that the transport industry as a whole competes with other need-satisfying activities for the allocation of scarce resources, the various forms or modes of transport, as well as transport enterprises, compete with one another for the resources which are indeed allocated to transport. The pursuit of maximum need satisfaction with the aid of limited sources requires that the scarce factors of production which are allocated to transport must be utilised optimally. Transport is therefore also subject to the economic principle and must satisfy the requirements for economic efficiency. Owing to the state's relatively significant involvement in the South African transport industry, it is difficult simply to use profit maximisation as the sole point of departure when investigating the behaviour of transport enterprises. However, we have already indicated that profit-maximising enterprises must also subscribe to the economic principle, and that the economic principle underlies the concept of economic efficiency. In the light of what has been said above, it can be stated that the field of study of Transport Economics embraces all those phenomena in the economy which influence the total opportunity cost of creating place and time utility by means of transport and its related activities. The economic principle serves as the point of departure in the investigation of the phenomena concerned. This comprehensive approach to transport economics means that both economic and business economic (or management) theory makes important inputs to transport economics as a subject, although such theory often has unique applications in the field of transport. The composition of the course on Transport Economics at the first-year level reflects this approach. In this module (TRL202Y), transport is considered mainly from the point of view of the community, and to a large extent, such module involves the application of micro-economic theory to transport. The other second-year module (TRL201X) is concerned with the management of transport enterprises. Before discussing the theoretical aspects of transport in more detail in study unit 4, we shall first consider the role of transport in a community.

SELF EVALUATION

(1) Distinguish between infrastructure and mobile elements of a transport system. (3)
(2) Define the economic concept of ***utility*** and indicate what two types of utility transport provide. (5)
(3) What does the derived nature of transport need mean? (2)
(4) Explain in detail the function and nature of transport. (20)

STUDY UNIT 3

The economic developmental role of transport

CONTENTS

3.1 THE ROLE OF TRANSPORT IN A COMMUNITY

3.1.1 General

The needs of a community involve far more than mere material welfare. Apart from the economic implications, access to various geographical areas and the mobility of people and goods also has specific social, political and strategic implications. In fact, transport is used by governments in order to achieve social, political and strategic objectives. In South Africa, the provision of subsidised transport for employees who, for former political reasons, stay in areas which are situated far from their places of work is an obvious example. In this section, we shall consider the economic role of transport in fairly great detail, after which we shall examine the social, strategic and political roles of transport.

3.1.2 The economic role of transport

3.1.2.1 Introduction

From an economic point of view, transport is important for two reasons. First, transport uses scarce resources in order to create place and time utility, and is therefore a scarce economic service which must be utilised optimally. This characteristic of transport constitutes the focal point of the present course and of Transport Economics, and will not be discussed further here. In section 2.1, we pointed out that the need for transport is a derived need, that is, that transport in fact serves a variety of activities, including economic activities such as production and consumption. The capacity of transport to provide, or increase, access and mobility by reducing the cost of bridging distances has a significant influence on the economy of a community. However, in order to understand this role of transport, it is necessary that we first consider briefly the historical development of market economics.

Thousands of years ago, people lived a nomadic existence and continually moved from place to place so that they could satisfy their basic needs by means of hunting and gathering. Gradually, however, they began to settle in places where they could plant and harvest crops, and where they could tend tamed animals. One of the consequences of this was that, while their farms produced surpluses of certain products, they experienced shortages of other foodstuffs. Owing to the surpluses which originated, some people were freed from having to produce food and began providing specialised services and goods. The surpluses produced were then exchanged for food. This allocation of different tasks to different people is termed "specialisation of labour". Specialisation of labour was far more efficient than the earlier subsistence economy because it allowed people to concentrate on those tasks which they could do best, and because, by concentrating on individual tasks, they could increase and expand their expertise and skills

considerably. In time, specialisation of labour enabled communities to provide for most of their needs by way of voluntary barter, which took place at a specific place or market.

This then was the beginning of what is today known as a market economy. With the origin of money, trading in products and services was facilitated considerably because it was no longer necessary for the parties to agree on the exchange of goods. Moreover, specialisation of labour was further promoted thereby. From a transport viewpoint, it is important to note that specialisation of labour made it essential for communities to trade in order to satisfy their needs. Technological development over the years and the increase in the demand for specific goods made large-scale production possible with the passage of time. This took specialisation of labour aimed at the provision of complete products by individual workers a step further, in that use was made of division of labour. This latter concept refers to specialisation in a specific activity (or part) in the provision of a product or service. The labour necessary to provide a product or service is divided into a number of successive and repetitive tasks, and individual workers concentrate only on the repetitive execution of a specific task. In order to enjoy the benefits of division of labour, large factories or other production facilities are necessary, which, in turn, requires the bringing together of a large number of workers and a large quantity of raw material as well as the large-scale distribution of outputs. Another important consequence of the entire process was the origin of product differentiation, which involves specific steps in the production process – ranging from the basic inputs to the production of the final product – being separated and then performed by individual enterprises, for example when the manufacturer of leather goods arranges for a separate enterprise to tan the leather. Specialisation and division of labour, as well as product differentiation, constitute the basis of modern production activities.

Owing to the fact that some communities, enterprises and individuals are capable of providing specific products and services at a lower cost than others, specialisation occurs not only within communities, but also at a regional, national and international level. Despite the greater production efficiency which is brought about by specialisation and division of labour, and by product differentiation, the impact or enjoyment of the full benefits thereof can be restricted by the lack of efficient transport. Not only does specialisation of labour and product differentiation necessitate the transportation of outputs, but division of labour and large-scale production also require effective access and mobility in order to make possible the supply of inputs and the distribution of outputs over great distances. The cost savings brought about by the more efficient organisation of production can be entirely or partly neutralised by the cost of transporting inputs from their sources or outputs to the markets where they must be sold.

Cole (1996:307) discusses the economic stimulus role played by transport. He states that developing countries' governments invest up to 40 per cent of total government spending in the transport sector. He concurs with the viewpoint of experts that transport is essential for development, but it cannot guarantee development automatically. A good transport infrastructure is a prerequisite, but not a guarantee. Other sectors like minerals, fuel, power production (electricity) and manufacturing should also be available. Transport investment should be balanced and at an optimum capacity to ensure development. However over-investment will have adverse effects on the economy and increase opportunity costs and lead to wastage of scarce resources. He concludes with this enlightening statement: ***"Transport facilities do not work economic miracles, and the integration of transport planning with other economic sectors is essential to the development strategy of a country."***

When the economic role of transport is viewed against this background, it is important to distinguish between the **passive and active** role played by transport. As the production and consumption activities of a community expand, the possibilities for specialisation and division of labour and for product differentiation increase, thereby creating a greater need for transport. In this regard, transport thus plays a **passive** role in that greater quantities are transported simply because the demand for transport has increased. In contrast, more efficient transport which is accompanied by a decrease in transport costs can create greater possibilities for specialisation and division of labour and for product differentiation. In this case, transport acts as a stimulus for economic development and thus plays an **active** role. Greater quantities are transported because transport costs have been lowered. A decrease in transport costs can have a significant influence on, among other things, the decentralisation of economic activities and the expansion of domestic and foreign trade. We shall now pay attention to specific aspects of the economic role of transport.

3.1.2.2 Transport and production

Transport may be viewed as an integral part of production or service provision. No production or service provision can occur unless the scarce resources or factors of production which serve as inputs are conveyed from their various origins at the right time to the place where production takes place. Furthermore, it may also be stated that the production or service-provision process is not complete until transport has added the necessary place and time utility to the particular outputs. What this all means is that transport costs can be regarded as constituting part of production cost. If transport costs decrease as a result of the more efficient provision of transport services, the factors of production can be supplied more cheaply to the producer. To this saving one can also add the decrease in the cost of transporting outputs to consumers, which means that the outputs can be supplied to consumers more cheaply. Alternatively, the decrease in transport costs can be employed to transport the inputs or outputs concerned over longer distances, or to transport more inputs or outputs over the same distance. In all three cases, production will increase to a greater or lesser extent, which makes the more efficient organisation of production by means of specialisation of

labour, mass production or product differentiation possible. A decrease in transport costs may also stimulate regional specialisation in production in that the comparative advantage which a specific region enjoys in respect of production costs (in the narrow sense of the word) can be utilised over a greater geographical area. The establishment of industries is also affected hereby. It can in fact be stated that transport promotes the more efficient organisation of production, helps to determine where production facilities should be established, and makes possible the large-scale concentration of the goods and people necessary to support a densely populated industrial area. As far as the distribution of outputs is concerned, efficient transport considerably increases the availability of goods and services by conveying outputs from various points to the market. Transport costs are an important consideration when taking distribution and marketing decisions.

3.1.2.3 Transport and the creation of utility

On a number of occasions we have mentioned that transport creates place and time utility. The transportation of goods from the place where they are manufactured to the market where they are offered for sale increases the value thereof to the consumer. The reason for this is that communities or individuals are placed in a position where they are able to acquire goods which they cannot readily produce themselves, or in respect of which they do not enjoy a comparative cost advantage as far as production is concerned, for example because of unfavourable climatic conditions or because the necessary raw materials, capital, skills or state of mind which make production on a competitive basis possible, is lacking. Such communities or individuals therefore experience shortages of the particular products, with the result that the provision thereof satisfies a definite need and is therefore of great value. Place utility is created directly by transport, while time utility is created either directly by transport or indirectly by means of storage or inventory holding. The direct creation of time utility by transport occurs especially in the case of perishable products which cannot be stored, in the case of time-sensitive products such as newspapers or stored data which cannot be transmitted cost efficiently by electronic means, and in cases where manufacturers do not hold inventories but use delivered materials directly for production, as in the case of just-in-time production systems. In such cases, reliance is usually placed on quick and reliable transport with a higher service frequency. However, the cost of rapid transport is mostly higher than that of slow transport. Moreover, a decision always has to be taken as to whether the higher cost is justified by the place and time utility which will be created. In a well-functioning market, the prices obtained will determine whether the additional transport costs are warranted. The transportation of an excessive quantity of products to markets where there is a shortage can force down prices to such an extent that the full transport cost cannot be covered thereby.

3.1.2.4 Transport and prices

The transportation of goods from places where there is an oversupply to places where there is a shortage should cause prices to rise in those markets in which the oversupply decreases and to decline in those markets in which the shortages are supplemented. Consequently, this will result in the equalisation of prices in the different markets. If a market is dependent on the fluctuating local supply of products or on irregular external supply, this can give rise to considerable price fluctuations between periods of good and poor supply. Efficient transport contributes to the stabilisation of prices by conveying products to other markets during periods of local oversupply and by bringing in products from outside during times of undersupply. Alternatively, it can ensure a regular inflow of products from outside. The equalisation and stabilisation of prices makes it possible for both producers and consumers to plan their activities better. The preceding arguments apply not only to consumer goods and services, but also to production factor markets, because remuneration for the factors of production in the form of, for example, profits, wages and rental is also stabilised and equalised.

3.1.2.5 Transport and competition

The supply of products, services and factors of production from different places to a specific market as a result of the availability of efficient transport will naturally give rise to increased competition among various suppliers in the market. Not only will this result in lower prices, but it will also contribute to improved quality and service provision. All this is made possible because a decrease in transport costs and a reduction in transport time dampen the influence of distance as a factor inhibiting entry to the market by different suppliers of products and services. However, the promotion of mass production by means of a decrease in transport costs may also give rise to monopolistic tendencies, since it may become more difficult for prospective competitors to enter the market on a scale large enough to be cost-competitive. The circumstances of each particular case will determine which one of the two contrasting trends will develop.

3.1.2.6 Transport and land-use patterns

Transport can influence land-use patterns in various ways. The creation of access to land and the mobilisation of goods and people in the vicinity thereof increase the land's production or utilisation potential for different possible uses, which are determined by factors apart from transport. Such factors include climate, topography, the availability of factors of production other than land, and historical land-use patterns. The increased production or utilisation potential of the land gives rise to new uses thereof, such as manufacturing. If improved transport reduces transport costs sufficiently, and if the specific environment lends itself thereto, regional specialisation and mass production can be promoted. The establishment of industries, regional production specialisation and the introduction of mass production in an area serve as important stimuli for urban development. Orderly urban development is promoted by efficient urban transport systems, including mass transport systems which enable commuters to reach urban destinations quickly and cheaply, and road transport systems which can accommodate both freight and passenger traffic efficiently. Furthermore, urban development requires efficient transport links with other cities and with rural areas. Transport's influence on location or land-use patterns also influences land values, since the opening up of areas for specific uses causes such values to increase considerably. It is therefore clear that transport can exert an important influence on economic growth and development.

3.2 THE SOCIAL ROLE OF TRANSPORT

Transport has a significant influence on the social structure of a community. The origin of large urban communities with ready access to social services in, among other things, the educational, health and cultural spheres would be impossible without the high degree of personal mobility brought about by modern transport systems. Furthermore, improved transport enables rural communities to share in useful social services and activities. Virtually every country in the world comprises different communities with divergent cultures, values and environments. Transport makes contact between different national and international communities possible, thereby promoting the interchange of ideas and values. As a result, prejudice decreases and uniformity in preferences, culture, education and general lifestyle is promoted. Reduced transport costs and time, and the consequent increase in personal mobility, can also contribute to the alleviation of serious socioeconomic problems. For example, unemployment can be reduced if cheaper transport is available to enable the unemployed to travel to areas where there is work. Transport can also make an important contribution in increasing the mobility of children, the aged and the handicapped in order that they may participate in essential and desirable activities.

3.3 THE POLITICAL ROLE OF TRANSPORT

The importance of a good transport system for an efficient system of government has been recognised for centuries (consider, for example, the roads which the Romans constructed in order to link the different parts of the Roman Empire with Rome). Authorities at different levels require effective access to those areas under their control in order that they may administer such areas efficiently and may provide those services for which they are responsible (such as health services in remote areas, ambulance services in both urban and rural areas, fire-fighting services, and the maintenance of law and order). In many cases, the authorities themselves have to establish and maintain the transport services necessary for the fulfilment of their responsibilities. One of the important political objectives of a government may be the integration of different communities by means of transport. However, political intervention in transport can detrimentally affect the effectiveness thereof. Examples of these are the establishment of a national airline in order to increase a country's prestige internationally when such action is not economically justifiable, and the provision of local transport facilities in order to gain political support, but which contribute little to satisfying the community's transport needs.

3.4 THE STRATEGIC ROLE OF TRANSPORT

In the case of military conflict, it may be necessary for a country to mobilise its resources over a wide geographical area and to concentrate them at specific places so that it can act effectively. Any government must, when formulating its transport policy, take into account the demands which may be made of the transport system under such circumstances, and must adapt the transport system accordingly. For example, specific components of the transport infrastructure (such as a road, harbour, airport or railway line) may, from a strategic military viewpoint, have to have a far greater capacity than would be required in peacetime.

SELF EVALUATION

(1) Briefly explain how the specialisation of labour and product differentiation led to the economic demand for transport. (10)

(2) Briefly explain the role of transport and production in the economic role of transport. (5)
(3) Briefly explain the role of transport and the creation of utility in the economic role of transport. (5)
(4) Briefly explain the role of transport and prices in the economic role of transport. (5)
(5) Briefly explain the role of transport and competition in the economic role of transport. (5)
(6) Briefly explain the role of transport and land-use patterns in the economic role of transport. (5)
(7) Fully discuss the economic role of transport. (30)
(8) Briefly explain the social, political and strategic roles of transport. (10)

PART 2

The transport market

OVERVIEW OF PART 2

The purpose of this part is to acquaint you, the student, with the transport market. In this part, you will really get to know what Transport Economics embraces, namely ascertaining how the transport market in which transport services are provided and utilised, functions. You will also begin to understand how the demand for transport differs from the demand for products, and how the demand for transport services is analysed. Similarly, you will discover the differences between the supply of transport and the supply of manufactured products. Cost of transport and the different cost types are of great importance in the supply of transport. You will also discover that there are differences between the costs which manufacturing concerns consider and transport costs. The fact that transport enterprises do not provide a single, homogeneous product, but a variety of services, will also become clear. The economic structure of transport is determined by the supply and demand characteristics of transport, and you will also gain a good idea of such structure. In addition, the transport market is distinguishable in that externalities such as pollution and congestion occur which are not included in the private costs of transport service providers. These costs must also be justified on economic grounds.

Lastly, we shall deal with price determination in the transport industry, where both costs and demand factors play a role.

STUDY UNIT 4

The demand for transport

CONTENTS

The aim of this study unit is to answer the following question: "What is the demand for transport?" Thus, in this study unit, we shall endeavour to place the demand for transport in perspective. Make sure that you are able to distinguish between

- the demand for transport services
- the demand for transport facilities

Transport economists, logisticians and entrepreneurs are confronted daily with having to solve problems in answering the transport question. In the past, engineers and transport planners, who were responsible for planning and designing the transport infrastructure, used models to analyse and predict transport demand, but paid scant attention to the economic content thereof. At this point, it is important that you become acquainted with the application of micro-economic theory to transport. It is also important that providers of transport services know what transport users are going to do, so that such providers can plan the transport services. Market demand, which comprises the total demands of all individual users, is in turn important as far as the demand for transport facilities such as roads, airports and harbours is concerned.

It is therefore important that you have a sound understanding of the demand for transport, and especially of how it differs from the demand for ordinary products and services. This will enable you to explain the interaction between transport demand and supply, so that you yourself can also solve transport problems.

For you to be able to explain clearly what the demand for transport embraces, you must be able to answer the following key questions:

- Why is there a need for transport?
- Who has a need for transport?
- How does the demand for transport differ from the demand for products?
- What does transport demand entail?

Key concepts

To enable you to understand why the demand for transport is important, it is necessary that we discuss the following core concepts:

- utility
- demand
- consumer surplus
- elasticity
- derived nature

Do activity 4.1.

ACTIVITY 4.1

Read the following paragraph and then answer the questions that follow.

On his farm, a farmer cultivates deciduous fruit for both the local and the export markets. Since the fruit is perishable, it has to be packed and transported to the nearest market on the day that is picked. Fruit is healthy and forms part of almost everyone's daily diet.

(1) Consider your own needs. Do you have a need for fruit?
(2) How do you get the fruit that you have in your home?
(3) Who has a need for the transportation of fruit?
(4) Why does the need for the transportation of fruit arise?
(5) By means of which form of transport will the fruit be transported?

Let us assume that you do have a need for fruit. Your need, and the needs of all those who eat fruit, will be known only to the farmer if you display a demand for fruit and the farmer is aware of such demand. You form part of the farmer's local market. Since you and the farmer are separated from each other by time and space, and you wish to eat his fruit before it perishes, both you and the farmer have a need for transport. This need for transport did not exist from the very beginning. Transport is necessary for you to be able to eat the fruit before it perishes. The demand for transport is therefore derived from the demand for the product at its destination (your greengrocer and eventually your home). This derived demand for transport is discussed fully in section 4.3.2. You should also realise that, because the fruit is perishable and therefore has to be transported quickly, a rapid form of transport, such as rapid road transport (refrigerated) or air transport, will have to be used. The specific choice will depend on the distance over which the fruit will have to be transported.

The above is a very common example of the role which transport economics plays in our lives. Transport economists are therefore interested in the amount of transport required at a particular time, in how much it will cost to satisfy such demand (supply), and in the prices which they must ask in order to cover costs. Remember that we are talking here about transport services and facilities. Thus the demand may, for example, be for a bus service (transport service), but it can also be for a highway or for a new harbour.

4.1 WHY IS THERE A NEED FOR TRANSPORT?

Before proceeding to answer the question posed in the above heading, we wish to refresh your memory regarding the concept "need" as it is viewed in the economy and in the economic sciences. This concept is central to all the economic sciences, and thus also to the transport economics which you are now studying. Economics considers the unlimited needs which exist and which must be satisfied by means of limited resources. In everyday life, there can be no demand for goods and services unless there is a prior need for it.

Do activity 4.2.

ACTIVITY 4.2

Assume that you are living in the desert. I approach you with an offer to supply you with sand at a very reasonable price. However, I am a person of many talents and am also prepared to supply you with an adequate quantity of drinking water but at a very high price.

Now answer the following two questions:

(1) Which of my two offers would you accept?
(2) Explain the reason for your choice.

I am sure that most of you (that is, those of you who do not stay near a permanent oasis) would rather purchase drinking water from me. Why? Because you need water in the desert, but not sand.

However, the problem remains one of how to get the water to you. I could ask you to collect the water from me, or I could convey the water to you in the desert. In both cases, transport, or a means of transport, is necessary. Without transport we cannot make use of the water. This is why we pose the question as to why there is a need for transport. Transport creates place (the desert) and time utility (thirst time). Transport was in no way what you ultimately needed. In fact, you were thirsty, but, to quench (satisfy) your thirst, you needed transport in order to be able to utilise (utility) the water.

The need for transport exists because there is a need for another product which can be satisfied only by making use of transport. Without transport, you cannot drink the water.

4.1.1 The transport need

What is meant by the concept "transport need"?

Above it was stated that the need for a product gives rise to a demand for such product. You also realised that you would have a demand for my water because you have a need for it in the desert. (I hope that you performed the above activity, otherwise you will not know what I am talking about now.) We will not know if there is any such thing as transport demand if we do not first look at the transport need. Why will there be something such as a transport need? I know that people need food, clothing and housing. Things such as food, clothing and shelter have utility for us and satisfy our needs. But transport does not provide us with essential utility. If I do not have a motor car, or do not travel by bus, train or taxi, or do not walk, I will still be able to exist. NOW can we say that transport has utility, that there is a transport need and that there is a transport demand? There is however, more correctly, a need for mobility, and transport is a means of providing such mobility. "Mobility" means that people (and goods) do not wish to be static and remain in one place, and the only way of bridging the distance between where they are and where they wish to be is by means of transport.

4.2 WHO HAS A NEED FOR TRANSPORT?

In Transport Economics we distinguish between goods transport and passenger transport. In section 4.1.1, we referred to the mobility of people and goods. People have a need for transport because they cannot perform all the activities which they want to at one particular place. Owing to the different land uses which occur in cities, people work, relax and make their purchases at different places. In order to move from one area to another, they need transport. Likewise, there is also a need for transport in rural areas. These needs are not for transport per se, but people wish to carry out their work, living, recreational and purchasing activities at different places and times. For this reason, passengers have a need for transport.

Products also have utility or value for consumers, but even products can only satisfy the needs of people if they are made available at the place where and time when they are required.

Transport is therefore required by the manufacturers of products in order that such products may be brought to the consumers of these products. Product manufacturers and distributors thus also have a need for transport, not for themselves as people or passengers, but for their products.

This distinction between the demand for goods and passenger transport is important for three reasons:

- It answers the question as to who has a need for transport.
- It helps us understand that the demand for transport cannot be viewed in only one way.
- It leads us to conclude that these different needs are in fact different parts of the transport need, which manifests itself as the demand for passenger transport or as the demand for goods transport.

The determination of who has a need for transport leads us to two strategies for considering transport demand. In this study unit, transport demand is of importance. Consequently, we must ascertain how a need for transport is converted into a demand for transport.

The need for transport is of importance because transport economists, planners, engineers and entrepreneurs cannot randomly provide transport facilities and services without knowing how great the need therefor is. There are two basic strategies for examining such need. These strategies are related to the economic system by means of which a community or country endeavours to solve the economic problem of satisfying unlimited needs with limited means. We may compare these two strategies to the capitalist free-market system as opposed to the centrally planned socialist system. In the first system, the interaction between market demand and supply and the price mechanism is employed as the point of departure for addressing the economic problem. In the second system, the socialist system's premise that the state can solve the economic problem by means of central planning and control is employed as the point of departure.

We can thus adopt one of two approaches in determining what the need for transport actually entails. First, we can hold the opinion that people need transport and that they must therefore be provided with it, regardless of whether or not they are able to pay for it. The state, which best represents the interests of its citizens, must then determine how great the need for transport is, and what means are required in order to provide for and satisfy such need. Transport is then viewed as a necessity and the state decides how the limited means are to be apportioned to people with transport needs.

Secondly, one can adopt the free-market strategy in respect of transport. In this case, the realised need as measured with reference to the market demand for transport is used. The demand for transport is then the amount of transport for which there is a need and for which users are also prepared to pay; in other words, the amount of transport which will be purchased in the market at any given time at different prices.

Hensher and Brewer (2000:72 and further) relate the need people have for transport to the **preference rules they follow and their budget constraints.** Also De Wit and Van Gent identify these influences on transport need and demand. All these authors state that both for passengers and goods there are certain attributes in the bundle of services offered which are of greater preference than others. Hensher et al therefore believe that shippers will choose truck over rail as the bundle of attributes is better for that consignment of the shipper. De Wit et al see these preferences for passenger transport relating to the age, educational level, sex or family composition of the people.

The budget considerations influencing travel need and demand relate to price of services, income and time.

Thus, when we speak of transport demand we mean the amount of transport which will be purchased in the market at any given time at different prices. Normally, the amount demanded will decrease as the prices increase, and, if a right-angled axis system is used to represent the relationship between price (along the vertical axis) and quantity (along the horizontal axis), the demand curve will slope downwards from left to right. It is important to distinguish between changes in demand and changes in quantity which are brought about as a result of changes in the prices asked. In the former case, the quantity demanded will change with each change in price and is represented by a shift in the demand curve, whereas, in the latter, the quantity demanded will move alongside the demand curve as the price changes.

If you think back to your studies of economics, you will realise that the above definition of transport demand accords with the concept demand as used in economics. It thus seems that we can apply the conventional demand theory of micro-economics to transport demand.

In order to apply the theory as decided above, we must answer certain questions concerning the key concepts indicated below. (The remainder of this section therefore contains a discussion of a number of basic concepts employed in demand theory.)

4.2.1 Utility and marginal utility

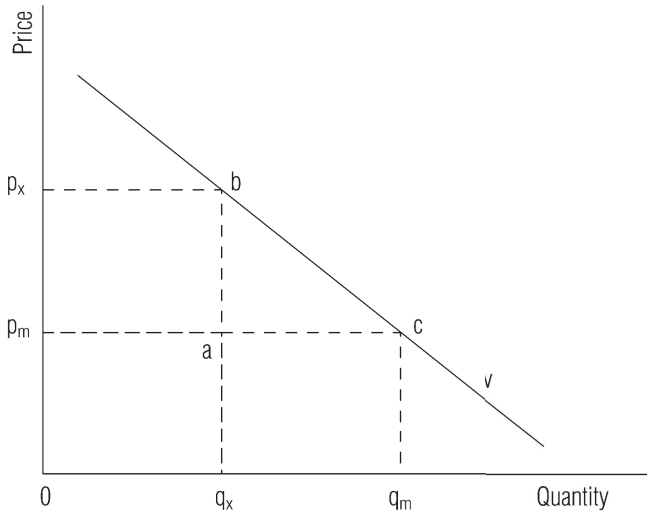
The concepts “utility” and “marginal utility” are of particular importance in explaining consumer or user behaviour. (In the literature on the subject, reference is usually made to transport users rather than transport consumers. However, the term “transport consumers” is increasingly being used . In this study guide, both terms will be used.) A person will purchase a product or service because it satisfies one or other of his or her needs; in other words, it has a specific utility for him or her. The magnitude of the need concerned determines the degree of utility which the product or service has for the buyer. Thus, the greater the need, the greater the utility of the product or service unit’s consumption. The converse is also true. The utility obtained from the consumption of a single additional product or service unit is termed “marginal utility”. According to economic theory, the marginal utility of a product or service decreases as greater quantities thereof are consumed in order to satisfy a need, the reason being that the intensity of the need decreases. This phenomenon is known as the “law of diminishing marginal utility” and may be formally defined as follows: The marginal utility which an individual consumer obtains from the consumption of an additional product or service unit diminishes as his or her total consumption increases. It is assumed that the individual consumer wishes to maximise the total utility which he or she obtains from the consumption of his or her limited means. It can be proved mathematically that the consumer obtains the maximum utility when he or she apportions his or her income in such a way that the marginal utilities per rand (or other monetary unit) of all the goods or services which he or she consumes are equal to one another. This result can also be derived intuitively. If the marginal utility per rand of the price of product A is greater than that of product B, the consumer can increase his or her total utility by consuming more of product A and less of product B. Consequently, the marginal utility per rand of product A will diminish whereas that of product B will increase, and the replacement of product B with product A will continue until the marginal utilities per rand of the two products are the same. This situation, which is termed “consumer equilibrium”, will obviously be disturbed if the consumer’s income, or the prices of the particular products or services, change.

4.2.2 The relationship between utility and demand

The individual demand curves can be derived with the aid of utility theory as discussed above. The point of departure is a condition of consumer equilibrium. If the price of product (or service) A declines, more of such product can be purchased per rand; hence the marginal utility thereof per rand will increase. The consumer will now purchase more of A and the marginal utility per rand thereof will diminish until equilibrium is restored. Normally, more units of a product are therefore sold if the price decreases. The converse can be derived in the same way. When this situation is represented graphically by way of a right-angled axis system, as in figure 4.1, it will be seen that the demand curve V slopes downward from left to right. The market demand curve assumes the same form as the individual demand curve, and is derived by summing the quantities which individual consumers demand at different prices.

FIGURE 4.1

The individual demand curve



Another economic concept which is frequently employed in transport economics, and which can be explained on the basis of utility theory, is that of consumer surplus. All the units of a product or service which a consumer purchases at the same time cost the same as the last or marginal unit which he or she purchases. However, smaller quantities, or the units which he or she purchased earlier, are of greater utility to him or her than the last unit which he or she purchases, and he or she therefore enjoys a "surplus" utility in respect of the units which precede the last unit. A consumer will continue to increase his or her purchases of a product until there is no longer a "surplus" utility per unit – which is quantity q_m in figure 4.1. For quantities smaller than q_m, the consumer will be prepared to pay more than the price p_m which he or she actually pays. For the quantity q_x, he or she will, for example, be prepared to pay a price p_x, which is ab higher than p_m. The surplus which the consumer enjoys is thus equal to the difference between the price which he or she actually pays and the price which he or she is prepared to pay. If therefore a consumer is faced with the prospect of the withdrawal of a particular product or service, he or she will rather be prepared to pay more for it. The total surplus of q_m which the buyer enjoys is equal to the area of the triangle between the demand curve and the straight line $p_m{}^c$. The phenomenon of consumer surplus plays an important role in public decision-making. Assume that a suburban road which can be used free of charge costs R100 000. In such a case, all the utility or benefits which users obtain therefrom constitute a surplus. If there are 1 000 road users, each of whom enjoys a surplus to the value of R100, they should bear the costs of constructing the road. If, however, the surplus of each is less than R100, they should not have the financial burden of constructing the road thrust upon them by way of the local authority.

4.2.3 Demand elasticity

We have already seen that the quantity of a product or service which is purchased in the market is sensitive to the price charged for such product or service. However, a criterion for measuring the degree of sensitivity must still be found. The elasticity concept provides such a criterion. Since price is not the only factor which determines the magnitude of demand, demand elasticity is

defined as the ratio of a proportional change in demand to the proportional change in the factor causing such change in demand. The price elasticity of demand (Ep) is for example expressed as:

$$\frac{\text{proportionate change in the quantity demanded}}{\text{proportionate change in price}}$$

$$= \frac{\triangle Q}{Q} \Big/ \frac{\triangle P}{\triangle P}$$

$$= \frac{\triangle Q}{\triangle P} \times \frac{P}{Q}$$

where $\triangle Q$ and $\triangle P$ represent the changes in demand and price respectively, and Q and P respectively represent the original demand and price. Apart from price, there are other factors in respect of which transport demand is sensitive – inter alia the income of the consumer, the prices of other products (cross elasticity) and the service level of the particular transport service. These factors are discussed in greater detail later on. Demand elasticity is usually expressed as percentages, and the importance thereof for the supplier of goods and services lays in the effect which these may have on his or her income. For the sake of convenience, the price elasticity of demand is used to illustrate this. The income generated from the sale of a product or service comprises a price component (P) and a quantity component (Q) and is equal to PxQ. Usually, changes in P and Q occur in opposite directions; in other words, if P increases, Q will decrease, and vice versa (as illustrated by the falling demand curve). Three cases may be distinguished:

4.2.3.1 Elastic demand

In this case, a specific percentage decrease in price causes a greater percentage increase in the quantity which is sold, which, in turn, means that the total income, namely PxQ, increases. In contrast, a specific percentage increase in price will result in a greater percentage decrease in sales, which means that the total income will decrease. In this instance, the price elasticity of demand is greater than one (Ep ≥ 4 1).

4.2.3.2 Unit elasticity

In this case, a specific percentage decrease/increase in price will result in an exactly compensatory percentage increase/decrease in the quantity demanded, and the total income (PxQ) will remain unchanged. This means that the price elasticity of demand is equal to one (Ep = 1).

4.2.3.3 Inelastic demand

In this case a specific percentage decrease in price causes such a small percentage increase in the quantity demanded that the total income (PxQ) decreases. Conversely, this means that a specific percentage increase in price will result in such a small percentage decrease in sales that total income (PxQ) will increase. This implies that the price elasticity of demand is smaller than one (Ep ≤ 1).

4.3 DIFFERENCES BETWEEN THE DEMAND FOR TRANSPORT AND THE DEMAND FOR PRODUCTS

4.3.1 General

Although there are certain obvious differences between transport and other user activities which require specific adjustments, micro-economic demand theory nevertheless provides a useful framework for analysing transport demand. The most important problems are usually those regarding the derived nature of transport demand, the fact that transport services cannot be stored for use at a later stage, and the specification and testing of appropriate demand models for travellers. The alternatives to a micro-economic framework are, inter alia, optimisation techniques, simulation and empirical analyses.

4.3.2 The derived nature of transport demand

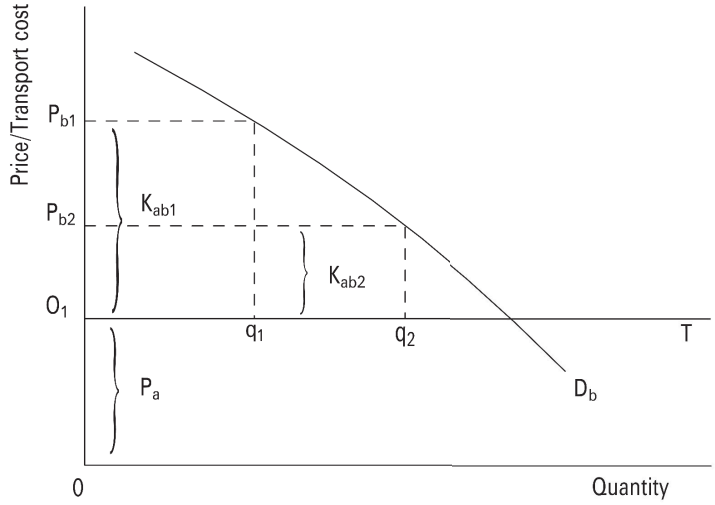
In very few cases is transport an end in itself. More often than not, the demand for transport originates from the demand for social and economic activities. Assume that a product must be transported from town A to town B. In figure 4.2 we indicate how the

demand curve for the transportation of a product from A to B can be derived directly from the demand for the particular product in town B. The curve Db represents the demand for the product at B. To convert this demand for the product into the demand for transport, one need rely merely on the fact that the price of the product at B (p_b) must be equal to the price of the product at point A (p_a), plus the transport costs between A and B (K_{ab}), that is, $p_b = p_a + K_{ab}$. If the horizontal axis in figure 4.2 is shifted upwards by a quantity p_a (which we assume is fixed), the curve D_b now indicates the quantity of transport (T) in respect of the new axis system which is demanded at different values of the transport cost (K_{ab}). At a transport cost of K_{ab2}, q_2 will be demanded.

Problems with the derived nature of the transport demand originate because travel and transport activities do not in themselves have utility (once again read through the discussion of the functions of transport in section 2.1), with the result that the maximisation of utility is not applicable. If, however, it is assumed that transport makes it possible to participate in social and economic activities which provide utility, and that transport is merely a means of overcoming the spatial separation between origin and destination, then transport cost can be viewed as an essential part of the total cost of consumption, or of whatever socio-economic activity is relevant. The consumer's utility must then be regarded as a function of the number of journeys which he or she undertakes and the number of associated socio-economic activities, while transport cost, or, more generally, the characteristics of the transport system which result in costs, can be included in the cost function or as a negative element in the utility function. From a theoretical viewpoint, there appears to be no reason why this adjustment cannot be made to the utility maximisation model. In practice, however, this is not an important issue because an ad hoc approach to determining the demand function (see section 4.1) is mostly adopted. In this regard, it is extremely important to distinguish between passenger and goods transport.

FIGURE 4.2

Derivation of transport demand



In the case of passenger transport, the utility function cannot be quantified and the ad hoc approach is generally used because the many peculiarities of individuals who take travel decisions serve merely to complicate any utility maximisation model and to make it unclear, especially as far as the combined market demand is concerned. This is why stochastic (probability) models are frequently used in order to determine the demand for passenger transport. In the case of freight transport, in contrast, transport may be viewed as part of the production process with inputs and outputs which are separated spatially. The optimisation of the production process provides a useful method of dealing with freight transport, and the utility function is indirectly quantified as part of the production function. This is in fact an important reason why passenger and freight transport should not be combined in a single, abstract transport demand, where transport users differ only in respect of the value of their user characteristics. Although doing so may be convenient for the analyst, it forces passenger demand to fit into the framework of production theory, which is suitable only for freight transport. The process by means of which utility is obtained from socio-economic activities, and, consequently, from the journeys associated therewith, is still not properly understood, but probably does not exhibit similarities to the production process, even given the fact that production decisions are also taken by individuals. In the case of passenger transport, the identification of the reason why a journey is undertaken (the objective of the journey) is an important means of stratifying demand

and of accommodating the derived nature of demand, while the same can be done in the case of freight transport by distinguishing between different commodities and production activities.

4.3.3 Why transport cannot be stored

In micro-economic theory it is assumed, quite correctly, that, in order to provide for fluctuations in demand, goods can be stored so that they can be supplied when there is once again a demand therefor. In the case of transport, this strategy cannot be employed to the same extent, and, in a certain sense, transport is provided even when there is not a demand for it, for example where an aircraft is ready to depart with empty seats or where a road is underutilised. However, it is not entirely true that transport cannot be stored. Some means of production which are used in order to provide transport services can in fact be stored until the necessary demand originates. As far as infrastructure such as airports and roads is concerned, fixed costs mount up regardless of the amount of traffic which uses such facilities. However, traffic-related variable costs need not mount up. Although a seat on an aircraft may be empty during a flight, the costs to the traveller, including his or her travelling time, are saved until he or she actually undertakes the journey by aircraft. The scheduling of transport services may be viewed as a method of storing such services and of offering them only at times when they are required. In practice, however, this amounts rather to the withholding of production.

4.3.4 Problems concerning traveller behaviour in the case of passenger transport

The most significant problems encountered in the application of micro-economic demand theory relate to the methodology and specification of appropriate demand models, and to the testing of the models concerned. The behaviour of transport users, especially as far as journeys are concerned, is subject to far more uncertainties than is the case with other consumption activities. In the case of urban transport, a typical urban household has an enormous amount of choices regarding possible combinations of location, number of journeys per day, modes, routes and departure times. The quantification of, and models for, all these choices and the associated decision-making process abound with assumptions, uncertainties and suppositions. Micro-economic demand theory therefore does not appear to be a suitable framework for the behaviour of a sociotechnical system such as transport. Consequently, a large number of stochastic models are encountered in the sphere of transport in which probabilities are employed in an endeavour to overcome lack of knowledge of urban travel behaviour on the part of transport planners. As regards experimentation, two factors in particular have an inhibiting effect. The first is the considerable extent of transport systems and the length of time which it takes for changes to work their way through such systems. By the time that the system achieves stability, the causative factors have already changed, thus making it very difficult to evaluate the consequences of the original changes. The second is that controlled experiments involving people are very difficult to conduct. In many cases, information concerning the behaviour of transport users in hypothetical situations simply amounts to speculation on the part of the respondent.

Two of the main analytical tools proposed by Hensher et al (2001:94–108) are linear regression models on the one hand and discrete choices and multinomial models on the other. Here mathematical models are used to use the preferences of people and their limited budgets to forecast transport demand.

In the demand for urban passenger transport these models are used to predict demand for passenger modes. Meyer and Straszheim in ***"Transport Economics: Selected Readings"*** by Oum et al (1995:7–13) explain how the concepts of **trip generation, trip distribution, modal choice and route assignment** lead to urban transport demand.

We can now look at the characteristics of transport demand.

4.4 CHARACTERISTICS OF THE DEMAND FOR TRANSPORT

4.4.1 General

In conclusion we shall consider the general characteristics of transport demand more closely. Bear in mind the micro-economic demand theory when studying this section, and endeavour to apply such theory where possible.

4.4.2 Variation in demand

One of the most outstanding characteristics of transport demand is the regular variation in the magnitude thereof at certain times. In urban areas, demand for road space and public transport systems is significantly higher in the early morning and late afternoon than it is during the rest of the day. In the case of intercity transport, the demand for passenger transport varies regularly during the year, with high seasonal peaks during holiday periods. Such peaks are a reflection of the variation in the demand for

socio-economic activities such as work, business, recreation and various social services which become available by means of transport. In the long term, economic cycles also cause variations in the demand for transport, with low demand during downturns and a lively demand during upturns.

Despite these variations, there is remarkable stability in the demand for transport. For example, households undertake the same average number of trips everyday, even though such trips may be made using different modes and for different reasons (journey objectives). Thus, more recreational and less work trips may be undertaken, or bus trips may be replaced by motor vehicle trips. This phenomenon can be the result of the fact that the travelling time available to transport users is possibly limited, especially where they wish to perform other activities at their destinations. However, research regarding travel time budgets reveals that the stable relationships indicated above may have a more complex explanation, thus requiring further research. Empirical investigations in the United Kingdom have shown that the average time spent on travelling increases in the course of time. The reasons for this are not clear, but it appears as if such phenomenon may be the result of rising incomes and the fact that each income group has its own fixed travel time budget. With the passage of time, many people have graduated from low-income groups with small travel time budgets to higher-income groups with larger travel time budgets. Research also indicates that travel time budgets increase more or less proportionately with income. These findings highlight the importance of time and other more conventional variables, such as income, for the analysis of transport demand (Button 1982:44).

4.4.3 The role played by the quality requirements of transport users in transport demand

"User preferences" is usually a collective term for all the variables (with the exception of the price of the product or service, the income of the user and the prices of other products and services) which determine the shape of the demand curve. Changes in user preferences therefore cause changes in the relationship between demand and the explanatory variables – with a consequent shift in the demand curve – rather than a movement along a demand curve which is based on established ratios.

The use preferences of transport users are definitely subject to change. The preference for private motor transport which has developed over time, as well as the shift in emphasis from price to quality of service, are examples of this. Both these changes can to some extent be attributed to increased standards of living. More money and free time have probably created a greater need for the freedom and flexibility associated with motor vehicle transport. Moreover, such a situation has probably given rise to other location patterns, with larger residential stands away from city nuclei, and, consequently, to different travel patterns.

The sensitivity of transport demand to service quality is a very important factor as far as user preferences are concerned. In the case of urban public transport, greater reliability, higher frequency and more bus shelters at bus stops are often preferred to lower bus fares. Furthermore, freight tariffs are not necessarily the most important determinant of the demand for freight transport. The results of an investigation carried out in the United Kingdom by the Price Commission are indicated in table 4.1. The special emphasis placed by shippers on the suitability of vehicles may be regarded as an indication of the value which they attach to factors such as protection against the elements, methods of safeguarding freight and the compatibility of the particular vehicle with its freight. The reasons why some shippers prefer to transport freight themselves at a higher cost correspond with the reasons why a particular carrier is preferred, and quality of service remains the most important consideration. This in no way detracts from the importance of price, but indicates merely that, in relation to quality of service, price decreases in importance as industrial production changes from the provision of raw materials and intermediary products to the provision of final and technologically refined products.

TABLE 4.1

Service characteristics which shippers expect of road carriers

Factor	Local (%)	Between regions (%)	National (%)
Suitability of vehicle	43	45	69
Rapid delivery	29	36	2
Rapid collection	10	12	14
"Good reputation"	15	5	1
Access to handling facilities	8	4	—
Condition of vehicles	—	—	8

Source: Button (1982:54)

4.4.4 Relativity of prices

Generally, it is accepted that the demand for a product or service is determined by the price of the particular product or service, by the prices of other products or services and by the income of the consumer. In the case of transport, however, the individual items require a number of refinements because, in reality, they are not simple variables, but are in fact combinations of interacting factors. The price of passenger transport does not, for example, include only travelling expenses but also all other costs, such as the time cost involved in the use of the service. Income is possibly not total income, but rather income above some or other subsistence level. Furthermore, clarity must be obtained regarding what precisely is demanded; in other words, is it merely a journey or is it possibly a journey by means of a specific mode along a specific route, or does it possibly comprise a "bundle" of transport services such as an aircraft flight together with the access journey to the airport? The individual demand factors will now be discussed in greater detail.

4.4.4.1 The price of transport services

The price of a transport service comprises far more than merely the travel fare or freight tariff which is directly paid for such service. For the purpose of modelling transport demand, and other quantitative work, the other price components are combined as an index of generalised cost (see section 6.6), which, among other things, includes travelling time, waiting time and transport safety.

It is difficult to generalise regarding the price elasticity of demand, especially where all modes are considered. However, it appears that a limited price change has comparatively little effect on the quantity demanded. Nevertheless, all factors must be taken into account. The demand for mass transport of raw materials by sea is, for example, inelastic because there are no substitutes therefor and because the demand for raw materials is comparatively inelastic. However, strong competition in the market may change this situation drastically. Investigations carried out in various countries have also indicated that the price elasticity of the demand for public mass transport is very low, with an elasticity of approximately -0,3 constituting the norm. (The value is negative because the change in price and in quantity usually occurs in opposite directions.)

As far as private motor vehicle transport is concerned, the effect of price changes on both motor vehicle ownership and motor vehicle usage must be taken into account. Investigations conducted in the United Kingdom have revealed that the elasticity of the demand for private motor vehicle transport is approximately -0,3 with regard to motor vehicle prices and approximately -0,1 with regard to fuel prices, whereas in the United States of America it is approximately -0,88 and -0,82 with regard to motor vehicle and fuel prices respectively. Investigations overseas have indicated that the price elasticity of the demand for air transport varies between -0,14 and -0,89 (Button 1982:45). The problem with general statistics on the price elasticity of transport demand is that the averages of a number of identifiable groups are calculated. The calculation of elasticity should be more specific, and therefore four broad criteria are distinguished for classifying transport users in groups:

TABLE 4.2

The price elasticity of fuel consumption

Investigators	Period	Short-term elasticity	Long-term elasticity
Houthakker et al (1974)	1963–72	—0,075	—0,24
Kennedy (1974)	1962–72	—0,465	—0,82
Reza & Spiro (1979)	1969–76	—0,210	—0,33

Source: Button (1982:48)

(1) Journey objective. There are clear indications that the demand elasticity as regards travel fares for journeys with specific objectives is far greater than it is for other journeys. Business trips, for example, are far less sensitive to changes in travel fares than recreational trips. The reason why a journey is undertaken must thus be taken into account.
(2) Method of payment. Users of different transport modes, and even of different services provided by the same mode, are frequently confronted by different methods of payment, and, consequently, their idea of the price paid for the journey may differ from the actual price thereof. Thus, for example, motorists take into account only a small portion of the cost of a journey by motor vehicle. They ignore, for example, the cost of oil, tyres, maintenance and use-related depreciation. In contrast, users

of public transport are far more aware of the total cost of their journeys because they have to buy a ticket. The use of season and travel tickets (bus passes) in such a case, however, tends to frustrate full cost observation.

(3) Passage of time. As in the case of any other purchase decision, the long- and short-term reactions of transport users to a change in price may differ considerably. Public transport users may react strongly to an increase in travel fares and, on principle, may make less use of such transport, but, over time, the intensity of their convictions may decrease. It is because of this that, in the long term, price elasticity is lower than the initial reaction suggested. Private motorists may react differently to increases in fuel prices, as is indicated in table 4.2 pertaining to investigations conducted in the United States of America. Initially, motorists showed little reaction to such increases, either because they regarded the change as temporary in nature or because they could not make changes in the short term owing to technical reasons. Similarly, commuting workers may initially show little reaction to price increases because they have to go to work, but, in the long term, the shifting of places of work and residence may result in significant changes.

(4) The absolute magnitude of price increases. Generally, elasticity increases as travel or transport distance increases. However, it must not be viewed merely as a function of distance, but in fact of price. A ten per cent increase on R5 is far smaller than a ten per cent increase on R500. Longer journeys are undertaken less often than shorter journeys, and then people gather more information on the former before making such journeys. In addition, such journeys are made for personal or recreational reasons; hence their greater elasticity.

4.4.4.2 Income levels

Although it appears that transport is a normal product in that the substitution and the income effect operate in the same direction when incomes rise, this is not true for all transport in all circumstances. Income has a positive effect on motor vehicle ownership, but as a result thereof has a negative effect on the use of public transport. In certain circumstances, public transport is thus an inferior product.

As in the case of price changes, the effect of income changes varies according to the duration of the period under review. In the short term, a decline in income may result in a significant decline in transport demand. In the long term, however, transport users may adapt their entire consumption pattern in order to be able to travel more, which means that the long-term elasticity of the transport demand is probably lower than the short-term elasticity.

The concept of a possible fixed travel costs budget is gaining increasing acceptance among transport economists. There are indications that households tend to spend a fixed portion of their income on transport. In the United Kingdom, a constant increase has been observed over the long term in the portion of income, or disposable income that is being spent on transport, while in the United States of America and Canada such portion has remained constant. Generally, however, it can be concluded that the view that there is some or other overarching budget mechanism which influences travel decisions is largely unproven.

4.4.4.3 The price of other transport services

The sensitivity of the demand for a product or service to changes in price in respect of other products or services is termed cross elasticity of demand. The cross elasticity of the demand for product A with regard to a change in the price of product B may be expressed as follows:

$$\frac{\text{proportionate change in the demand for A}}{\text{proportionate change in the price of B}}$$

$$= \frac{\triangle Q_A}{Q_A} \Big/ \frac{\triangle P_B}{P_B}$$

$$= \frac{\triangle Q_A}{\triangle P_B} \times \frac{P_B}{Q_A}$$

The demand for any particular transport service will be influenced by the actions of competing or complementary suppliers in the sphere of prices, or, strictly speaking, of all other suppliers. The relationship between motor vehicle costs and the demand for public transport has already been indicated. There is also a crosswise effect among the different public transport modes. An interesting phenomenon is however the almost total insensitivity of the demand for urban motor vehicle transport to both bus and train fares. This is probably the reason why attempts by local authorities to restrict the use of motor vehicle transport by subsidising public transport have been largely unsuccessful.

Information on the cross elasticity of the demand for complementary transport services, for example feeder services to long-haul services, is very limited. The expansion of through-road networks undoubtedly increases the demand for transport on certain feeder routes by lowering motor transport costs on such roads, while at the same time resulting in declining demand on competing routes. The exact implications of such changes are far more difficult to determine than is, for example, the case with changes in mode division, but they are most definitely an important characteristic of transport systems.

4.5 CONCLUSION

By now you should know why there is a need for transport, who experiences a need for transport, how transport demand differs from the demand for other products, and what transport demand is.

You should also now be able to apply conventional demand theory, with the necessary adjustments, to transport. If you understand the similarities between transport demand and conventional demand theory, and take into account the differences between transport and other products, you should be able to describe the principal characteristics of transport demand.

SELF EVALUATION

Answer the following questions on transport demand.

(1) Define the following concepts:

- utility
- marginal utility
- consumer surplus
- elasticity
- need
- transport demand

(2) Why is there a need for transport?
(3) What two groups of users have a need for transport? Explain in full why we distinguish between these two groups.
(4) Explain in detail the three differences between transport demand and the demand for products.
(5) Describe the principal characteristics of transport demand.

Q1 five marks for each dot. (5)
Q2 five marks (5)
Q3 ten marks (5)
Q4 & Q5 thirty marks each (30)

Although you now know what the demand for transport is and why it is important that transport economists (and others) are able to determine transport demand, we have discussed only one aspect of the matter. In order for you to have a proper understanding of the transport market, it is necessary that we consider the supply of transport. In the next study unit we therefore deal with the supply of transport.

STUDY UNIT 5

The supply of transport

CONTENTS

In the previous study unit, we looked at "one side of the coin". In both micro-economics and transport economics, demand and supply are of cardinal importance. In this study unit, therefore, we shall endeavour to place transport supply in perspective. Make sure that you are able to distinguish between

- the supply of transport services
- the supply of transport facilities

The reason why you must be able to distinguish between the two is that the supplier is not a single, identifiable entity. Transport services are supplied by a wide variety of entities, while the transport infrastructure is provided by the state or a government institution. Furthermore, the transport user plays a greater role in the determination of transport supply than in the case of the supply of manufactured products.

The aim of this study unit is to acquaint you with the distinctive nature of transport supply. The entire transport system serves as the point of departure, and the role of the transport user in the origin of transport supply also receives attention. After studying this study unit, you should be able to discuss and explain the following: The framework for analysing transport supply, the factors determining transport supply, the participants in the origin of transport, and the interaction between such participants. You must also be able to determine the components of the supply function, and to indicate the principal differences between transport supply and the supply of manufactured products. To be able to explain with insight what the supply of transport entails, you must be able to answer the following key questions:

- What is meant by "the supply of a product or service"?
- What variances occur in the sphere of transport which necessitate that we consider transport supply differently from the supply of manufactured products?
- How does transport supply differ from the supply of manufactured products?
- What is transport supply?

5.1 INTRODUCTION

The supply of a product or service may be defined as the quantities of the product or service which producers are prepared to offer in the market at different prices. This is usually represented graphically by a curve on a right-angled axis system, which, as in the case of the demand curve, indicates the relationship between market prices (along the vertical axis) and the quantities offered at each price (along the horizontal axis). Since the quantity offered normally increases as the price increases, the supply curve

rises from left to right. The sensitivity of supply to market prices is known as the price elasticity of supply, and it is formulated in a similar way to the price elasticity of demand. Elasticity does not play such an important role in the case of supply as it does in the case of demand. Generally, the supply of transport, especially as far as infrastructure is concerned, is relatively inelastic. There is not necessarily a direct relationship between the producer's production costs and the quantity of a product or service which he or she is prepared to sell at different prices. In a market which is competitive in respect of both the necessary factors of production and the products or services which are provided therewith, production costs do have a significant effect on prices to the extent that they affect the producer's decision to produce. To continue with production in the long term, the price which the producer or supplier receives must at least cover his or her costs, otherwise he or she will have to discontinue production and will no longer be able to offer his or her products on the market. The role played by prices in transport is discussed in-depth in study unit 9.

The above definition of supply can be accepted in the case of products or services where the price of the producer or supplier is the principal determinant of the quantity demanded at the particular price. In the case of transport, however, a number of adjustments need to be made to such definition. One of the most important reasons for this is that, apart from the price which he or she pays, the consumer also has to make valuable inputs to the transport process. The magnitude of the additional costs is dependent on the service characteristics and level of service of the particular transport system, and is, in fact, also influenced by the travel and transport behaviour of the user. In the case of transport, there is a clear relationship between demand and supply. The remainder of this study unit will be devoted mainly to the complex process by means of which transport supply originates.

5.2 FRAMEWORK FOR THE ANALYSIS OF TRANSPORT SUPPLY

The problems encountered in applying micro-economic supply theory to transport are more serious in the case of combined market supply than they are in the case of supply by individual producers. Generally, however, there are three variances in the sphere of transport which require an amendment of the conventional definition of supply.

(1) The supplier or producer is normally not a single, identifiable entity whose behaviour can be investigated explicitly. In the case of urban transport, the system by means of which supply occurs is so comprehensive that it is, for example, impossible to identify a single supplier who can determine what the total supply should be in terms of prevailing prices. Where no direct payment (toll) is levied from users, such users pay indirectly for the use of the transport system by means of income, municipal, fuel and other taxes. These problems decrease where a single supplier is responsible for the provision of infrastructure, control systems and the final transport service, as happens in the case of national railroad enterprises.

(2) As already discussed, there are a number of nonmonetary aspects of supply, especially service quality, which are just as important as, and sometimes more important than, the price which the transport user must pay. For example, travelling time is a characteristic of transport supply which can have significant cost implications for the transport user. The conventional approach to supply must thus be adapted in order to take into account the different characteristics of transport supply which have cost implications for the transport user.

(3) In many respects, it is the behaviour of the transport user rather than that of the supplier which determines the characteristics of transport supply. Many of the major factors which determine the service level of a transport system are dependent on the way in which transport users utilise the existing transport system, and are therefore not determined by the supplier. In the case of road transport, the road user's route selection has, for example, a significant effect on his or her travelling or transport time, while his or her choice of speed can also have an important influence on travelling or transport time, as well as on the running costs of his or her vehicle. It is particularly when total transport demand is analysed, or when projections are made with regard to traffic volumes, that all the characteristics of transport influencing the quantity and nature of transport activities which do in fact occur, have to be taken into account. The complex interaction between the behaviour of suppliers and users is directly and indirectly responsible for the use of scarce economic sources in the transport system, although they cannot always be quantified or expressed in monetary terms. The selection of the appropriate characteristics which must be taken into account depends on the nature of the particular type of transport. In the case of urban motor vehicle transport, travelling time, running costs, traffic delays and the availability and cost of parking are possibly sufficient to describe transport supply, while in the case of intercity air transport it may be necessary to take into account travelling time, air fares, service frequency, the flight schedule, the loading factor, ground costs, the type of aircraft, the level of passenger service and delays at the airport. It is by no means necessary to make use of a single definition of supply for all types of transport. A single framework would probably be complex and would probably confuse important aspects of transport analysis.

5.3 DIFFERENCES BETWEEN THE SUPPLY OF TRANSPORT AND OF MANUFACTURED PRODUCTS

5.3.1 The storing of transport services

This matter was also touched upon in the previous study unit. It is not possible to keep the final output of a transport enterprise, namely transport services, in stock for sale at a later stage if such services are not sold immediately. This limitation has a number of important implications. First, marketing experiments are more risky than is the case with manufactured products. If, for example, a price increase is too great, unsold stock of manufactured goods can be sold later at reduced prices in order to minimise losses. In the case of transport, part of the market may be alienated in such a case, with very costly consequences. The possible loss of revenue and custom is consequently of greater importance in the case of transport enterprises than it is in the case of manufacturing enterprises, and the former will endeavour to avoid marketing risks.

Second, the existence of unsold capacity close to departure time may force the transport operator to take extraordinary measures to obtain custom, for example by offering discounts on freight or travel tariffs and by taking other direct measures to entice customers away from competitors. The continuing existence of unutilised capacity in the long term may result in the adoption of a more strategic approach in order to attract customers. Conventional price encouragement, and even the offering of unconventional services and the development of non-traditional markets, may form part of the strategy. In cases where both the demand and supply of transport services are inelastic in respect of prices, price decreases may give rise to a decline in the revenue of transport enterprises, which may, in turn, result in further price decreases. In such circumstances, the likelihood of uneconomic price competition becomes very strong. The nature of competition in the transport market is dealt with in greater detail in study unit 7.

5.3.2 The flexibility of production

In most cases, manufacturing enterprises are able to deal with changes in their demand patterns more easily than transport enterprises. The ability of the former to store production makes it possible for them to increase their inventory holding when demand is low, and, when it is high, to reduce such inventory holding. This allows for a constant output level and plant utilisation, with market demand still being satisfied. However, at any specific moment, the saleable output of a transport enterprise is all that can be sold in order to satisfy demand, because no saleable stock is available. Any change in output as regards sales necessarily implies a change in capacity utilisation or an expansion in capacity, both of which have significant cost implications. Apart from inventory holding, manufacturers can resort to other ways of changing their available output so as to adjust to demand. In particular, they can increase their output in the short term by way of more intensive utilisation of fixed factors of production and the concomitant greater consumption of variable factors of production. Similarly, production can be cut back by delaying the production process and thus reducing variable costs.

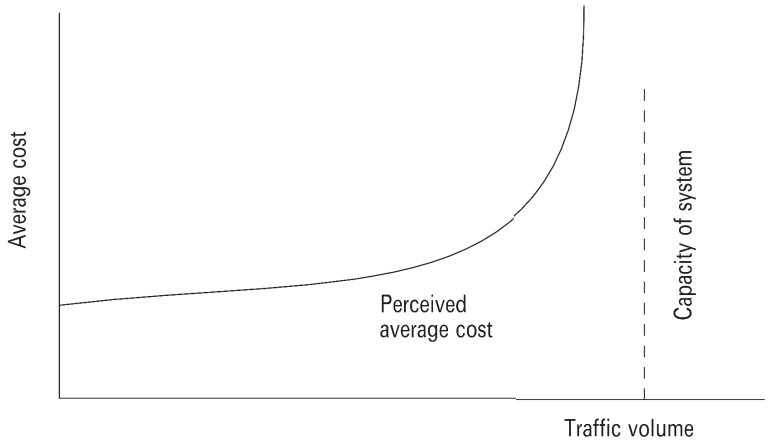
The service frequency demanded limits the short-term elasticity of supply, especially in the case of scheduled transport services. A decrease in service frequency means a lowering of the service level, which may not only increase the observed costs of the user, but may also detrimentally affect the user's image of the service. Other service level requirements can also complicate capacity adjustments considerably. Passengers and freight shippers want transport at the moment when they need it. Although transport supply is relatively inelastic, the supplier of transport services must in some way or another cope with the cyclical, non-recurring or random changes to which transport demand is subject in order to retain the goodwill of his or her customers. In practice this means that there must always be sufficient capacity to deal with regular and unexpected fluctuations in demand. Such capacity is usually greater than the average demand and increases the average costs of transport supply. From an operating viewpoint, this means that a capacity amount which is greater than the average demand must be available in order to maintain the level of service. This contingency or reserve capacity is therefore by no means unnecessary, and may be the only practical way of satisfying peak demand. The long, useful life of transport infrastructure and some means of transport means that reserve capacity, or any surplus capacity, cannot be reduced in the short term.

5.4 THE SUPPLY FUNCTION

In the light of the above discussion, it is now possible to provide a clearer definition of transport supply. Transport supply, as represented by the transport supply function, is an expression of the relationship between the service level characteristics of the transport system as observed by users and the output volume of the system. It is thus in fact an expression of the inverse of the causal relationship involved in the demand function (as defined in study unit 4). The demand function indicates how traffic volume is influenced by the service level characteristics of the transport system, and the supply function indicates how these characteristics are influenced by the volume of the system's traffic which is used.

FIGURE 5.1

Transport supply function



Source: Kanafani (1982:65)

If all the characteristics of the transport system which determine the service level of such system are viewed as components of transport cost, the supply function is similar to the function of the user's average total cost, because the average cost of each traffic unit determines what volume of traffic will materialise, as indicated in figure 5.1. For this conversion to be meaningful, it is important to use those costs on which the user actually bases his or her use decisions, namely the perceived cost of the transport concerned. The supply function then represents the relationship between traffic volume and perceived cost. This requirement is of particular importance when the costs of individual motor vehicle users must be included in the supply function, for it is fairly common for individual vehicle owners to underestimate the costs related to the use of their vehicles. Perceived cost also plays an important role in the compilation and use of the supply function of scheduled public bus, air and rail transport. Passengers consider waiting time to be more inconvenient than the time which they actually spend travelling by means of the various modes. Consequently, the perceived cost of waiting time is higher than that of transport (or travelling).

As a rule, the difference between perceived and actual costs is determined by way of empirical studies. Demand-and-supply models based on empirical studies usually represent user behaviour, and thus, by implication, the user's observation of system characteristics. According to economic theory, the market will be in equilibrium when demand and supply are equal, that is, where the demand and supply curves intersect, or where the demand and supply functions are equal. As far as the latter condition is concerned, an equilibrium analysis can be conducted far more easily if the same variables are used to describe traffic volumes and service level in both functions. If both functions are converted to cost functions by making use of perceived costs, the equilibrium point will indicate the volume of traffic which will result at the particular level of perceived cost.

5.5 THE COMPONENTS OF THE SUPPLY FUNCTION

5.5.1 General

To be able to determine the supply function, the service characteristics of the transport system which entail costs for the user, and which are consequently determinative as regards the traffic volumes which are realised, must be identified. In this regard, it is necessary to distinguish between passenger and freight transport, because there are important differences between the cost components of their supply functions.

5.5.2 Components of generalised transport cost

5.5.2.1 Passenger transport

In determining the supply function concerned, the following components are usually taken into account in the case of passenger transport:

(a) Total travelling time

Total travelling time can be divided into the following components: access time (ie the time which the traveller takes to move from the origin of his or her journey to the departure point of the mode concerned, eg an airport), waiting time at the departure point, transfer time (where the traveller has to transfer from one means of transport to another) and the time actually spent on travelling. Not all these components may be relevant or of equal importance.

(b) Total travel costs

This component is also subdivided into a number of subcomponents, including the following: Direct costs, vehicle running costs, indirect taxes, and terminal costs in the case of certain modes, such as parking costs for private motor vehicles.

(c) Inconvenient departure times

Scheduled services such as bus, air and rail transport services do not always have departure times which exactly match the times at which users wish to depart, with the result that the user may be inconvenienced. Such inconvenience can be measured in terms of the amount of time which elapses between services, but such approach is valid only until each traveller or passenger's desired departure time is known. Usually the time between a traveller's desired departure time and the scheduled departure time is regarded as part of waiting time, and therefore of total travelling time. However, it is more appropriate to view such time separately, because travellers do not necessarily wait from the moment of the desired departure time, but instead depart later from their origins.

(d) Travel comfort and convenience

It is easy to recognise and identify the effect which this characteristic has on traffic generation, but it is by no means easy to measure it. In some cases it should, however, be included in the supply function. In analysing the supply of air transport, it is for example accepted that, if all other factors remain the same, the greater comfort of jet transport entails a lower average user cost than non-jet transport.

5.5.2.2 Freight transport

The components which are taken into account in the case of freight transport correspond in many respects with those taken into account with regard to passenger transport, but the way in which the transport user observes them, and the consequent effect thereof on the supply function, naturally differ. The following cost components are usually taken into account:

(a) Total transport time

Total transport time is divided into long-haul time and terminal time. The latter occurs at both the origin and the destination and includes the time taken to load and unload freight, to transport such freight to and from the terminal, as well as any delays which may occur.

(b) Total transport cost

Total transport cost is divided into line-haul cost, terminal cost (including transportation to and from terminals) and storage cost at the terminals concerned. In some cases, line-haul tariffs also cover other cost items such as insurance and taxes. When the transport user and the operator are one and the same entity, operating costs, insurance and all taxes become directly identifiable components of the total transport costs of the user.

(c) Service frequency

As in the case of passenger transport, an increase in frequency results in a decrease in transport costs for the user. A service with a higher frequency results inter alia in a decrease in inventory holding, and consequently in the storage costs of freight shippers.

(d) System reliability

This component has to do with the ability to predict the performance of the transport system, such as arrival times and the condition in which the freight will be delivered. Such capability increases as system reliability increases, and may also result in a decrease in inventory holding because smaller buffer stocks are required as system reliability increases. The reliability of a system can be quantified by measuring the variation in time, cost and safety.

(e) Preservation

Preservation is a very important characteristic in the case of goods which can be easily damaged or spoilt, and relates to the deterioration of freight during the transportation process. The rate at which goods perish will depend on the method of packaging or transportation. Fruit which is transported in refrigerated vehicles perishes more slowly than when it is conveyed in unrefrigerated vehicles. Cooling can thus reduce the total transport costs involved. Although a high level of service, as represented by inter

alia a high frequency, the availability of adequate capacity and high system reliability, can reduce the total transport costs of the transport user, it can also considerably increase the costs of the supplier or operator. A high frequency may require a greater number of vehicles, thus increasing operating costs; adequate capacity in order to provide a service on demand may involve the creation of reserve capacity; and greater reliability may result in increased maintenance costs and may require the provision of stand-by vehicles and equipment. The higher costs of maintaining a higher level of service will naturally result in increased prices, which the user will have to weigh up against the advantages of a higher level of service when he or she chooses a transport service.

5.6 FACTORS DETERMINING THE CHARACTERISTICS OF TRANSPORT SUPPLY

5.6.1 Characteristics of transport supply

The particular way in which the characteristics of transport supply manifest themselves depends to a large extent on the transport technology employed, on the operating strategy adopted by the operator, on institutional requirements and constraints, and on the behaviour of transport users.

5.6.1.1 Technology

Transport technology is described mainly in terms of support (the vertical contact between the vehicle and the surface on, or medium in, which it moves), guiding (the steering of vehicles or vessels by means of reaction forces), propulsion (the type of drive unit and the method of transfer of acceleration and retardation forces), control (the means used to regulate the movement of one or more vehicles in a transport system) and the body and bodywork of transport units (which fulfil a containerisation function). (Refer to the study guide for TRL201X for a more detailed discussion.) From an overall point of view, the most important type of technology is that related to road, rail, pipeline, air and sea transport, although subdivisions are encountered within each transport form. The technical characteristics of a transport system have a significant effect on both the performance and cost thereof. In particular, the relationship between the fixed and variable costs of transport depends to a large extent on the type of technology employed. Important characteristics of supply, especially those related to service quality, such as speed and capacity, are dependent on the technology used.

5.6.1.2 Operating strategy

The objectives of an operator determine his or her behaviour and the way in which the technology of his or her choice is utilised in order to provide a particular transport service. The strategy which is for example employed in order to provide a specific capacity, or to expand capacity, is determinative as regards several of the supply characteristics of scheduled systems such as air and bus transport. Larger aircraft require a lower service frequency than smaller aircraft in order to provide the same capacity. The behaviour of the operator will also determine how much of his or her costs he or she will recover from transport users, as well as the way in which he or she will recover such costs. By means of the particular pricing mechanism which the operator employs, he or she converts his or her production or operating costs into user costs. The role played by prices in the transport industry will be discussed at a later stage.

5.6.1.3 Institutional requirements and constraints

The operating strategy or pricing policy of transport operators is frequently subject to the requirements and constraints of the regulatory system or prevailing market conditions under which they operate. The regulatory system may control the capacity which the operator offers by way of operating authorisations or permits, may fix and control his or her tariffs by means of administrative or hearing procedures, and may impose certain transport obligations on him or her. Furthermore, the type of vehicles and equipment, as well as the way in which they are operated, may be prescribed by transport and traffic legislation. The market structure, or the nature of competition in the transport market, can also have a significant effect on the behaviour of operators. A transport monopolist may, for example, limit his or her supply in order to keep tariffs high, and may also apply price discrimination. By means of collusion and the formation of cartels it is possible for an oligopolistic market to function like a monopolistic market, or to be very unstable, with general tariff wars accompanied by wastage. In a competitive market, in contrast, operators are price takers and cannot influence the market by means of collusion. Consequently, they have to adapt their operating and price strategies to the needs and prescriptions of the transport market. Market structure and transport competition are dealt with in study unit 7.

5.6.1.4 User behaviour

At various stages in the foregoing discussion we indicated the effect which the behaviour of transport users can have on the characteristics of transport supply. In particular, we pointed out the effect which their choices in respect of aspects such as routes, speed and mode can have on the transport supply to urban travellers. The behaviour of freight shippers also has a significant effect on the characteristics of the transport supply to them. The way in which freight shippers use the available transport facilities and services frequently determines their total transport costs: Shippers can inter alia vary their inventory levels, the sizes and frequency of their consignments, and their packaging methods. By ordering small consignments at short intervals a shipper can lower his or her inventory holding, but will at the same time increase his or her transport costs. Moreover, a shipper's packaging methods will determine his or her handling requirements and time, and consequently his or her handling costs. The interaction of the above four factors in bringing about transport supply is illustrated in figure 5.2.

5.6.2 Participants in the supply of transport

In order to facilitate the description of transport supply, it is useful to describe the four most important participants in the process as well as the contribution of each.

5.6.2.1 The supplier (or producer)

Here we are referring to the person or entity who or which provides a transport service. Such person or entity may be a roads department which is responsible for constructing and maintaining a road system, or an enterprise which possesses a number of vehicles or other transport means which are used in order to provide a transport service. The operating behaviour of the supplier, as influenced by prevailing market conditions or the operation of the regulatory system, influences the particular transport technology concerned to achieve a specific level of technological performance. Technological performance as well as the supplier's behaviour influences his or her costs directly. The supplier can also act as an operator or regulator. In many countries, the planning, construction, operation and regulation of airports are undertaken by three different government bodies. If the supplier and operator are one and the same entity, the supplier recovers his or her costs directly from the transport user, or, if other participants are involved in the operating or price determination process, such costs are recovered indirectly. Spoornet, the rail section of Transnet, is for example both the supplier and the operator of its infrastructure and rolling stock, and therefore recovers its provision costs directly from its users. In contrast, the provision costs of roads departments are recovered indirectly by means of taxes and levies, or, in some cases, by means of toll fees collected by operators.

5.6.2.2 The operator

The operator is the entity responsible for the operation of the transport system as well as for daily decision making regarding particular service characteristics, such as scheduling and routing. Apart from recovering his or her operating costs by means of tariffs, the operator is also sometimes involved in the recovery of the supplier's costs by means of direct levies. The cost recovery system which is used converts the costs of the supplier and the operator into user costs. In the case of some transport systems, the supplier and the operator are one and the same entity, but often they are separate entities which jointly offer a transport service, with the supplier providing the basic infrastructure and the operator assuming responsibility for the operation and maintenance of the system. This applies to both public and private transport. The supplier or operator need also not be a single entity. In the case of air services, the basic infrastructure may be provided by the state, with the operating components of the system, such as aircraft and other equipment, being provided by air transport enterprises. Furthermore, the airport may be operated by a local authority, with the air transport enterprise operating its own aircraft. As far as private motor vehicle transport is concerned, the road infrastructure is provided by government institutions, with the motor vehicles and their accessories being provided by the operators, who are at the same time also the users.

5.6.2.3 The transport user

A transport user is the person who or entity which takes the travel or transport decision. In the case of travel decisions, such person is the traveller or passenger, and in the case of transport decisions it is the consignee or dispatcher of freight or the consignor (shipper). The transport user is important in the analysis of supply not only because his or her behaviour influences the service level of the transport system, but also because the transport costs as observed by him or her have a significant effect on the supply function. Since the observed costs frequently differ from the actual costs, and because travel and transport decisions are based on observed costs, the appropriate adjustments must be made when converting supplier and operator costs to user costs. The same applies to all aspects of the level of service which affect the travel or transport decision: That which is observed by the

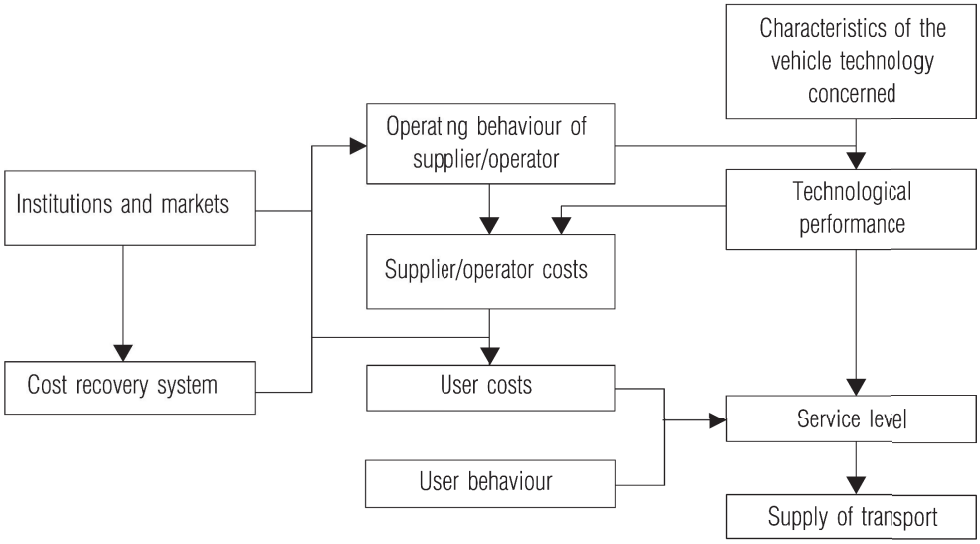
user is important in analysing supply. Such matter is further complicated by the fact that the user's behaviour often has an effect on actual costs and other aspects of the level of service. This is why the supply function which is used in a systems context for demand analysis is specifically defined as a combination of supplier and operator costs and their methods of cost recovery, as well as the effect of user behaviour on these costs and the user's observation thereof.

5.6.2.4 The regulator

The regulator is the entity which exercises some or other form of control over the operation of the transport system or over the relationship between the participants in the supply of transport. Regulation may be either technical or economic in nature. Regulation of a technical or operational nature relates to the requirements regarding the way in which transport technology must be adapted, provided or operated. Economic regulation relates mostly to the methods of cost coverage and price determination, as well as to the quantity and nature of transport services which are provided. Regulation can affect all three of the above-mentioned participants in the supply of transport. In the case of the construction of an airport, the relevant government body (a supplier) must adhere to specific norms or standards, while the operator who operates the airport as well as aircraft operators must comply with the prescribed safety regulations. Both operators must adhere to specific rules regarding the way in which clients must pay for their services. A functional, hierarchical relationship exists between the four participants in the supply of transport. This relationship begins with the supplier, whose choice of technology essentially determines which type of transport will be offered. Thereafter, there is the operator, who adapts and manipulates the technology in reaction to traffic conditions and the environment in which the system is operated. Finally, there is the user, who is the recipient of the transport service. The regulator has an effect on all three of the aforementioned participants. This functional hierarchy also exists in the case of costs and level of service. The supplier's costs influence those of the operator, and the operator's costs, in turn, have an effect on the user's costs. The costs of all three, as well as the relationship between such costs, will be influenced by the regulator.

FIGURE 5.2

The generating of transport supply



Source: Adapted from Kanafani (1983:60)

5.7 CONCLUSION

In this study unit, the supply of transport was dealt with mainly from the point of view of the total transport system. Furthermore, it became apparent that the conventional definition of supply in micro-economic theory has to be adapted in order to make provision for the distinctive characteristics of transport supply. In particular, the behaviour of the transport user must be taken into account. The amount demanded in the market is determined by the service level characteristics of the system, and is reflected in

the amount (volume) of the system's traffic which is used. The supply of transport is in fact nothing other than the relationship between service level (which is the counterpart of price in conventional demand theory) and the volume of output of the transport system (which is the counterpart of the supply of products in conventional theory). However, you must constantly bear in mind that it is the user's observed costs of the service level characteristics of the transport system which determine his or her use thereof. Although particular emphasis has been placed in this study unit on transport supply in a systems context, the same principles apply to a greater or lesser extent to the supply of transport by separate facilities in the systems or by individual operators.

SELF EVALUATION

(1) Why can conventional supply theory not be applied unaltered to transport? (10)
(2) Are there any significant differences between the supply of transport and the supply of manufactured goods? (20)
(3) Discuss the most important factors that determine the characteristics of transport supply. (10)
(4) Explain, with the aid of a diagram, how transport supply arises as a result of the interaction of the factors that are indicated in question 3 above. (8)
(5) Define the demand and supply of transport and compare it to the conventional definitions of demand and supply in micro-economic theory. (10)
(6) Explain what is meant by the equilibrium between demand and supply in the transport system. (5)
(7) Discuss the conversion of the service characteristics of the transport system to user costs. (30)

STUDY UNIT 6

Transport costs

CONTENTS

OBJECTIVES

After having studied this study unit, you should be able to:

- describe the nature and characteristics of economic costs
- define the relevant cost concepts with reference to the avoidability of costs
- explain fully the behaviour of total and unit costs in both the short and long term
- understand the multiproduct nature of transport services
- explain the implications thereof on cost behaviour

In the previous study unit you learned what transport supply is. You saw that such supply differs from the supply of manufactured products, and that one must always bear these unique characteristics of transport in mind when evaluating transport supply.

Think back a little. What are the three variations which are encountered in transport which make transport supply unique?

Think of a practical service such as passenger commuter services between the suburbs of Pretoria and the city centre.

(1) The railway line and infrastructure are owned by Spoornet. The services are provided by Metrorail, and the Commuter Corporation is the operator which sells the tickets. There are thus a number of entities involved and not only one.
(2) Not only the cost of the train ticket, but also aspects such as how regularly the train operates, which stations will be stopped at, how comfortable the seats are, and (especially in South Africa) the personal safety of passengers will play a role. Thus service quality and other nonmonetary aspects are important.
(3) Whether the passenger is an auditor or a clerk will determine whether he or she will travel to work by train or by motor vehicle. This will also determine where and when rail commuter services will be offered. The commuter or transport user contributes to the determination of supply. Consequently, it is not only suppliers who determine the supply on their own.

You now know how transport demand and supply are interlinked. But the cost curves constitute the basis of the supply curve. Which costs are of importance as far as transport is concerned? Remember that we are considering economic costs. In this module, we shall also be considering costs from an economic and not a management point of view. Costs from a management viewpoint are examined in another module, namely Transport Management TRL201X. You must bear this in mind throughout. It is particularly cost behaviour within the transport industry which is relevant rather than how costs and cost control occur within a transport enterprise.

To enable you to discuss transport costs with insight, you must be able to answer the following questions:

- What are economic cost?
- What is meant by the following cost concepts: Direct cost, indirect cost, joint cost, common cost, total cost, fixed cost, variable cost, marginal cost and average cost?
- How do costs behave (cost behaviour)?
- How does cost behaviour in a multiproduct enterprise (in other words, a transport enterprise) differ from that in a single-product enterprise?

The aim of this study unit is to provide you with knowledge of the nature, identification, structure, and behaviour of economic costs so that you will be able to understand aspects of transport economics such as the economic structure of transport and price determination with regard to transport.

6.1 INTRODUCTION

The availability of objective and realistic information on the economic costs of transport services is a prerequisite for effective decision making by all participants in the transport market. Decisions concerning inter alia tariffs, the expansion or reduction of capacity, and the choice of a transport service require reliable cost information. The principles underlying the identification and measurement of costs are relatively simple, but their application gives rise to several complex problems. In this study unit, we shall pay attention mainly to the private economic costs of transport provision or transport operation as seen from the viewpoint of the supplier or operator. At the end of the study unit, the costs of the transport user are also considered more closely. The external costs of transport which, together with the private costs, represent the total costs of transport for the community are discussed in study unit 8.

6.2 THE NATURE OF ECONOMIC COSTS

I recently inherited R40 000 from my grandfather. I must now decide what to do with this amount of money. I can either invest such sum in a bank or I can use the money to start my own transport business. I am considering buying a friend's kombi so that, in the mornings, I can transport children in the vicinity to nursery schools. What would the opportunity costs be if I were to buy such vehicle? What must I consider if I wish to take economic costs into account? Does the fact that my own vehicle also cost R40 000 three years ago have anything to do with the economic costs?

My opportunity costs with regard to the R40 000 constitute the interest which I would lose by buying the kombi and not investing the money. My old motor vehicle has absolutely nothing to do with the matter – it represents a sunk cost which cannot now be recovered.

The above example serves to illustrate that opportunity cost is in fact the cost of an alternative opportunity of which I could make use. We shall now explain opportunity or economic costs in more detail.

The factors of production which are used directly or indirectly in order to provide goods or services are scarce in relation to all the possible uses thereof. As a result of the fact that the demand for goods and services which can be provided with the relevant factors of production exceeds the supply thereof, the factors of production possess a specific value. The magnitude of the value of the factors of production is determined by the availability thereof for the production of goods and services on the one hand, and by the demand of producers therefore on the other. If the factors of production are scarce in relation to the demand therefore, this means that the use thereof in order to provide a particular product or service will prevent such factors being used to provide other products and services for which there is also a demand.

This argument constitutes the foundation of the concept of economic cost, namely that the cost of using factors of production to provide a product or service is equal to the value of the particular factors for the production of the next best alternative product or service. This cost is known as the opportunity cost of the particular factors of production. Such concept applies from the highest level, where decisions have to be taken regarding the utilisation of factors of production in different sectors of the economy, to the

level of the individual enterprise, where decisions have to be taken regarding the utilisation of specific means of production. To determine the economic or opportunity costs of factors of production can be a very complex process. If the particular factors of production are traded in a well-functioning market, the price which is arrived at by consensus in the market serves as a measure of the opportunity costs thereof. In cases where the market does not function well, or where there is no market for the particular factors of production or their products, or where government action affects market prices, market prices will no longer reflect the scarcity value or opportunity cost of the production factors, or no market prices will be established. In such cases, prices must be estimated or market prices must be adjusted, and shadow prices are therefore used (see study unit 10).

Conventionally, the private or internal costs of a supplier of transport facilities or of a transport operator comprise the payments which he or she makes in order to obtain the factors of production which are necessary for the provision of transport facilities or transport services. One of the principal characteristics of opportunity costs is that they have a bearing on the future. Obviously, factors of production which have been consumed in the past are not available for alternative uses. Moreover, the historical cost thereof does not form part of the economic cost of future production or service provision, even if there are still financial obligations relating thereto. Sunk cost also does not form part of the cost of future production. Some production activities require such specialised means of production that they cannot be adapted for alternative use, for example railway and road tunnels or special handling equipment. Sunk cost is equal to the difference between the opportunity cost, which is avoidable before a decision is taken, and the value which can be recovered once a decision has in fact been taken. Sunk cost does not form part of the opportunity cost of future production, but the recoverable value of the means of production still makes provision for alternative uses and thus forms part of opportunity costs. Only the costs associated with the use of means of production with alternative uses, or those with some or other recoverable value, thus form part of the economic or opportunity cost of future production or service provision.

This is one of the most important reasons why economic and financial costs do not always correspond. In order to be able to allocate, either wholly or in part, the costs associated with the use of specific production factors to the provision of a specific output, a causal relationship must exist between the particular costs and output. A method which is commonly used to relate costs to output is to ascertain whether the input costs can be avoided by not producing the particular output. Only the costs associated with the production factors which are wholly or partly used to produce the output can be completely or partly avoided. The identification of avoidable costs is in fact very important, because these are the only costs which can be changed by means of decision making. Avoidable costs can be identified at several levels, from the reduction in the supply of a single product or service to the closure of an entire enterprise or system. The concept of avoidable costs is useful in facilitating the explanation of certain other cost concepts.

6.3 GENERAL COST CONCEPTS

De Wit et al (1996:68) indicate that the classification of costs into different categories depends on the goal of the enterprise or institution. The different uses of different cost categories will determine how the costs will be categorised. If the costs are categorised by a transport enterprise according to their operations; costs can be operational and non-operational. Operational costs can then be subdivided into direct and indirect costs.

For the explanation of how different cost categories react over time and use of vehicles we have fixed and variable costs. If we are interested in costs per unit we have average or marginal costs. If the cost analyst is interested in the allocation of costs to the production units (different vehicles or cost centres) we distinguish between common and joint costs. The concept of user costs relate to the experience of the transport user. Especially in passenger transport, the waiting times, convenience or inconvenience, comfort, walking times etc all play a role in the total perceived user costs. User costs are important to the transport supplier as well as users who do not only base their decisions of transport usage on the production costs, but on their perceived user costs.

In studying these cost concepts you should keep this in the back of your mind.

6.3.1 Total, average and marginal costs

The total cost of a transport service comprises all the payments which are made in order to obtain the means of production necessary to provide a specific quantity of the service. Such cost is therefore the accumulated monetary amount required for the entire output of a transport service.

The average cost of the service is obtained by dividing the total cost by the number of units of the service which are provided.

The marginal cost of the service is the total cost associated with a small increase or decrease in output, usually a single unit. The relationship between average and marginal cost is essentially a mathematical rather than an economic one. The most important aspect of this relationship is that, when marginal cost is less than average cost, average cost is in the process of decreasing, and that, when marginal cost is greater than average cost, average cost is in the process of increasing. The increase in total cost as a result of an increase in production or service provision is also termed incremental cost, and the saving in total cost as a result of

a decrease in production or service provision is termed avoidable cost. Although the terms marginal, incremental and avoidable costs are frequently used synonymously, marginal cost is, by definition, the cost associated with a very small increase or decrease in production, while incremental or avoidable costs refer to the costs associated with greater changes in production or operations. The term "avoidable costs" is however often used instead of "incremental costs" in that a possible increase in total cost as a result of service or production expansion is regarded as the cost which can be avoided by not taking such a decision.

6.3.2 Fixed and variable costs

By now you should be familiar with the concepts of fixed and variable costs. Fixed costs are costs of which the magnitude does not change as output increases or decreases within the limits of existing capacity, even where the output decreases to zero. The costs of constructing a road or railway line, for example, remain unchanged, but the road and railway line will have to be maintained, even if they carry no traffic whatsoever. The same applies in respect of the cost of owning vehicles and other means of transport. Fixed costs can be avoided only to the extent to which they are not sunk costs. In contrast, variable costs vary as the magnitude of output changes and can be avoided by producing no output. Apart from the fixed costs associated with the construction and maintenance of terminal facilities, total terminal costs increase as the amount of traffic which is handled increases. Variable costs thus constitute that part of total costs which can be avoided if output falls to zero. Average variable cost is the variable cost per output unit for a specific output level.

The distinction between fixed and variable costs becomes important when one has to determine the avoidable costs of an output unit (a service unit or transport performance), that is, the variable costs of transport performance which can be avoided by not rendering such performance. As will become clearer later on, avoidable costs represent the theoretically ascertainable costs of transport performance. The rendering of transport performance may also require the use of durable means of production such as terminal facilities, but as long as the technical or design capacity thereof is not exceeded, the more intensive use of durable means of production will not increase the costs thereof. However, the replacement of a durable means of production such as an aircraft or a truck is not without cost, and the replacement cost thereof must be written off against the total output which such means produces during its useful life. As far as avoidability is concerned, this means that such cost is causally related to the accumulated production of the durable means of production and not to the individual transport performance which is rendered therewith. The fact that fixed costs are in most cases indivisible (ie cannot be divided among specific output units) does not mean that variable costs necessarily have to be divisible. Variable cost differs from fixed cost only in that it originates as a result of the use of nondurable means of production. If more than one output unit is produced with a unit of a nondurable means of production, the variable cost of the units thus produced is indivisible. It is for this reason that the choice of units in terms of which output is measured is important. The smaller the unit selected, the greater the magnitude of indivisible costs in respect thereof. To avoid the problems associated with the indivisibility of costs, output units must be selected (bearing in mind technical considerations) in such a way that the greatest cost amount is divisible.

Normally, a transport service involves a large number of output units or transport performances which is much smaller than the total output of the vehicles and other means of transport which are used. Although it may be appealing to regard ton kilometres or passenger kilometres as the relevant output units, it is more meaningful to use the return journey of a train, bus, motor vehicle, truck, ship or aeroplane as the relevant output unit. The margin between fixed and variable, or avoidable, costs is largely dependent on the period under review. The longer the period, the greater the portion of total cost which is variable or avoidable. In the short term, very few costs will change as the output of a transport enterprise or transport system changes. All decisions have already been taken and all means of production, with the possible exception of energy and labour, have already been allocated. Certain output-related costs, such as the use-related maintenance costs of terminal facilities and equipment, will on the face of it appear to be fixed, but will materialise at a later stage. In the short term, there are thus very few avoidable costs.

As the period under review becomes longer, fewer and fewer decisions will be fixed and increasingly less sources will be allocated to specific outputs. In the medium term, vehicle, vessel or aircraft fleets can be adapted to changed market needs, and the associated operating strategy can be changed. In the long term, it is possible to expand, reduce or change fixed facilities such as airports and roads in accordance with changed needs. In the short to medium term, the capacity of a bridge, for example, is fixed, but in the long term it can be widened and reinforced in order to handle a sustained increase in traffic, or it can be allowed to deteriorate and not be rebuilt if the volume of traffic decreases or is expected to cease entirely. As in the case of the divisibility of costs, the magnitude of variable costs is also dependent on the output unit which is used as a criterion. Rail freight transport can be used as an example. If a train which is about to depart has an empty truck available, the variable or avoidable costs of an additional truck load (or less) comprise the additional energy, labour and administrative costs. Where a complete transport service is introduced or withdrawn between two places, a portion of the asset cost, such as depreciation and interest on capital, as well as a portion of the tractive force cost are included in the variable or avoidable costs. At a still higher level, the cost of a transport service which is offered over the entire railroad network, for example a coal transport service, or the operating cost of an entire link or greater part of the network, may be regarded as variable or avoidable. At the highest level, a decision regarding

the continued existence of an entire transport enterprise or transport system is necessary, and then the total cost of the enterprise or system will be variable or avoidable.

6.3.3 Common and joint costs

The concepts of common and joint costs have more to do with cost apportionment than with a specific type of cost. In order to be able to fix tariffs at a later stage, for example, all costs must be apportioned to the outputs which are produced. The determination of the causal relationship between costs and output is hampered by the nature of transport output. In the case of any transport enterprise, different types of traffic are handled between different origin-and-destination pairs at different times. Similarly, any transport system handles a variety of traffic types. The same equipment and facilities are therefore used in order to provide a variety of transport services. In this regard, two cases can be distinguished, namely where the costs associated with shared means of production are common, and where they are joint. We shall first define these cost concepts before discussing them any further.

6.3.3.1 Joint costs

Joint costs are the costs which originate as a result of the absolutely associated production of two or more products.

Joint costs occur when, owing to technical reasons, the use of the means of production to produce one output necessarily leads to the production of a second output in a fixed ratio to the first output. The most obvious example of this is the forward and return journey of a vehicle. If a vehicle which is stationed at A is used to transport a load from A to B, it must of necessity return to A. The avoidable cost of the forward journey is obviously the total variable cost of the entire return journey, for this cost originates even if the return journey is an empty-leg journey. From an economic viewpoint, the transportation of freight during the return journey entails no cost (except, possibly, for higher fuel and labour costs), because the cost of such journey exists in any event and can thus not be avoided. The same argument applies where a forward journey is undertaken so that freight can be transported on the return journey. From a theoretical point of view, avoidability does not provide a measure for apportioning the costs of a return journey to forward and return loads. From a practical viewpoint, the solution is to allocate joint costs to the load which is transported in both directions. The return journey is thus regarded as an output unit. If freight is transported in only one direction, the joint costs must be allocated thereto, and, if freight is transported in both directions, the joint costs must be allocated to the combined forward and return loads.

6.3.3.2 Common costs

Common costs are those costs which originate when two or more products are manufactured in one production establishment or enterprise without there being any technical connection. Such costs are thus divisible in principle.

The costs associated with shared means of production are common when there is not necessarily a fixed relationship between the different outputs which are produced therewith. The portion of the means of production which is used to produce one output could just have well have been used to produce more of the other, and vice versa. In the case of rail transport, the same track, signalling system and shunting facilities are often used for a variety of freight and passenger traffic, and the ratio between the quantities of each service which is provided can be changed relatively easily. The occurrence of common costs decreases as the output unit which is used as a criterion becomes greater. The cost of tractive power is common to the different truck loads of a train, but can be directly allocated to the train itself. The fact that the occurrence of common costs makes it impossible to determine the causal relationship between cost and output does not prevent the apportionment of an identifiable part of common costs to a specific output unit. For the purpose of apportionment, it is useful to regard a transport enterprise as a collection of intermediate processes and activities which, in combination, produce the final output which is sold to users.

6.3.4 Direct and indirect costs

The apportionment of costs is also relevant in the case of direct and indirect costs.

Direct costs are those costs which can be directly apportioned to a product, whereas indirect costs are costs which cannot be apportioned directly to a specific product. It is especially the multiproduct nature of transport services which makes cost apportionment such a problem.

6.4 COST BEHAVIOUR

6.4.1 General

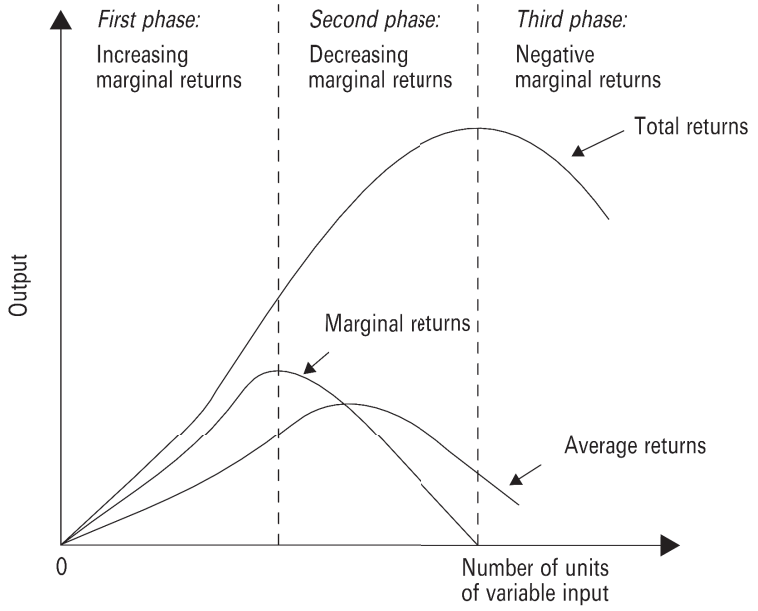
Generally, the relationship between the fixed and variable costs of an enterprise or industry are referred to as the enterprise's cost structure. However, the provision of products and services occurs under a wide variety of technical, market, management and financial conditions. These conditions or factors can, either individually or collectively, result in enterprises which essentially produce the same product differing significantly as regards size, cost structure and cost level. These differences are related to the behaviour of costs when the ratio between factor inputs, or the size of a plant or enterprise, changes. To enable you to better understand these phenomena, we shall first discuss a number of important topics relating to production theory.

6.4.2 The law of diminishing marginal returns

In the production or operations process a large variety of production means are combined in order to produce the planned output or outputs. When the ratios between combinations of production means are changed, the magnitude of the output which is produced therewith usually also varies. This phenomenon is the result of the operation of the law of diminishing marginal returns. According to this law, when increasing quantities of a variable production factor are combined with a fixed production factor, the total, average and marginal returns will initially increase, but a point will be reached where the increase in returns of the variable production factor will be proportionately smaller than the increase in output. It is even possible for the marginal returns to decrease to zero. It is important to note that this law applies for a given state of technology and of factor prices, and that it has a bearing only on the relationship between production factors, and not on their absolute magnitude.

FIGURE 6.1

The operation of diminishing marginal returns



The curves in figure 6.1 indicate the change in returns (output) as the input of the variable production factor is increased while the other production factors remain constant. Initially, the total, average and marginal returns increase more than proportionately as more units of the variable factor are combined with the fixed production factors. However, a stage is reached when, first, the marginal returns per unit of the variable production factor begin to decrease, which, in due course, also results in a decrease in the average returns per unit of the variable production factor, at which point the total returns begin to increase less than proportionately in relation to the increase in variable input. When the marginal returns reach zero, the total returns at the same time reach a

maximum, and, when the marginal returns become negative, the total returns will begin to decrease. From an economic point of view, the second phase, as indicated in figure 6.1, is the most important. A producer or operator will increase his or her input of the variable production factor in the first phase, because, by doing so, he or she will be able to achieve a more than proportionate increase in his or her output. In contrast, he or she will not take his or her output past the second phase because his or her total output will then begin to decrease. It therefore follows that, from an economic viewpoint, production or operations should take place only in the second phase, in which diminishing marginal returns occur.

Furthermore, this means that, at any specific moment, economic constraints are placed on the total output of a plant or enterprise, and these constraints are determined by the course of diminishing marginal returns. The basic reason for the change in output owing to the operation of diminishing marginal returns is the change in the ratio of production factors used. The choice of a specific ratio between production factors will determine in which phase of diminishing marginal returns goods or services will be produced. The producer's choice will be determined inter alia by the relative cost of the particular factors. The most economic combination of production factors at a specific output level is that which results in the lowest total cost per unit (average total cost) of the particular service or product.

As already mentioned, the optimal ratio between fixed and variable production factors depends on their respective prices. If the price or cost of the fixed production factor is high relative to the cost of the variable production factor, it will be advantageous to expand production by adding more units of the variable production factor until the marginal returns thereof are very low. This is possible because the decrease in average fixed cost is sufficient to compensate for the increase in average variable cost, which follows on the increase in marginal cost. Marginal cost is equal to the cost of an additional unit of the variable production factor divided by the marginal returns thereof. If the marginal returns decrease, marginal cost will thus increase. Consequently, average total cost still decreases for significant increases in output from the point where the marginal returns began to decrease until the minimum average total cost is reached. These conditions usually occur where a fixed factor of production is provided in large indivisible units and possesses considerable unutilised capacity. The higher the cost of such a fixed production factor, the more profitable the more intensive utilisation thereof will be. If, however, the cost of the variable factor of production is high in relation to the cost of the fixed production factor, the expansion of production will probably no longer be profitable once the marginal returns of the variable production factor equal the average returns thereof. The reason for this is that the increases in marginal and average variable costs after this point are too large to be counteracted by the decrease in average fixed costs, and the average total cost will consequently begin to increase rapidly up to the point where marginal and average returns begin to decrease. After this point, there is thus a rapid movement away from the optimal combination of production factors.

6.4.3 Economies of scale

6.4.3.1 The nature of economies of scale

It is important to distinguish between diminishing marginal returns and economies of scale. The course of diminishing marginal returns determines the optimal output level of an enterprise under given conditions, but provides no criterion for the economically optimal, absolute size of a plant or enterprise, even where the state of technology is accepted as given, because it has a bearing only on ratios. The economic size of a plant or enterprise is determined by scale returns, and, more specifically, economies of scale. The concept of economies of scale refers to the greater efficiency, or lower unit cost of production or operations, which is achieved by the expansion of the plant or enterprise. Initially, scale increases were defined as increases in production capacity without any change in the ratio between production factors, and thus, by implication, without any changes in the products or service provided and in the technology employed. Essentially, this amounted to an enlarged version of the smaller-scale unit. More recently, however, such strict requirements have no longer been explicitly laid down. However, when the unit costs of enterprises are compared to different production capacities, it is important that such comparison is confined to plants which produce products or product lines which are similar enough to be regarded as competing in the market.

Furthermore, the production processes must correspond, particularly as regards the amount of processing which occurs within the plant. The scale of a transport enterprise which undertakes both freight consolidation and long-haul transport cannot be compared to that of enterprises which are involved in only one or two processes. The scale of plants which use different technology can also not be compared to one another, for example rail or bus transport. To realise economies of scale, the scope of the market must be sufficient to be able to absorb the increased production. An enterprise's production can be adapted so as to serve a small market at the lowest possible cost, but an increase in the scope of the market may make it possible for such enterprise to expand the scale of its production so as to realise possible economies of scale. A transport operator may, for example, use a larger bus, aeroplane, ship or truck in order to reduce his or her unit cost, provided that the variable costs are not too high.

One of the most important reasons which is usually given for economies of scale is the indivisibility of especially fixed factors of production; and one of the most important examples of this is the provision of the fixed track (road or rail), which, for technical

reasons, can be done only in large indivisible units. In other cases, it is merely possible for plants or equipment with larger capacities to achieve a lower cost per unit of a product or service than smaller plants or equipment. Many vehicles fall into this group. Indivisibility is also applicable in respect of labour and management: A larger plant or operation can often be handled by the same management team and work force. It is sometimes argued that the better utilisation of management is the greatest source of economies of scale. This naturally implies a change in the ratio between production factors, which is contrary to the initially strict definition of scale expansions, which required that the ratio between the input quantities of production factors remain unchanged. Another important source of economies of scale is specialisation. A higher output frequently makes it possible to organise the production or operations process more efficiently by way of an increased division of labour and specialisation.

It is not only the tasks of labourers which lend themselves to greater division and specialisation, for such division and specialisation can also assume the form of more specialised machinery, instruments and automatic control systems, as well as more efficient methods of material handling, maintenance and inventory holding. In addition, production planning and production control systems can be better integrated. Scale expansions do not result only in the reorganisation of tasks with a view to achieving greater benefits from specialisation in production or operation activities, but also in changes in the skills and scope of the work force, as well as in changes in product technology and in the ratio of labour to capital.

6.4.3.2 Scale returns

Economies of scale are realised in one of the three scale return phases in which an enterprise's production plant may find itself, namely in the phase of increasing scale returns. In this phase, output increases more rapidly than the inputs of production factors. This phase is restricted inter alia by the scope of indivisibilities and by the exhaustion of task division and specialisation. A very important additional constraint is that relating to the complexity of management. After increasing scale returns have been exhausted, there follows a phase of constant scale returns. In this phase, the different conflicting forces are in equilibrium, and output increases in the same proportion as production factors. It would appear that a large number of enterprises are in this phase. In the third phase, the increase in output is proportionately smaller than the increase in factor inputs, and consequently diminishing scale returns are experienced, although total output may still increase.

6.4.4 Large-scale production

The concept "large-scale production" can have various meanings which, although comprising common elements, may differ considerably. Large-scale production is, for example, not necessarily associated with a large enterprise. A controlling company may control a large number of enterprises which produce different outputs and which are each fairly small. The controlling company's size is however determined by the number of enterprises which it controls, not one of which need have at its disposal the benefits of large-scale production. "Large-scale production" generally refers to the difference in size between giant enterprises (such as steel and vehicle manufacturers, and enterprises which provide rail transport and generate power) and the many small enterprises which are to be found in the agricultural and building sectors, in numerous service sectors and in the retail trade. The term may also be used to refer to the scale of production within a specific industry, such as large-scale farming as opposed to small-scale farming. However, the most commonly accepted meaning includes both large-scale production or operations and large-scale organisation, and is therefore related to economies of scale in respect of production and enterprise size.

6.4.5 Other sources of savings

Economies of scale are not the only source of lower unit costs in the transport industry. **Economies of density** and economies of scope may also occur. Economies of density is a concept which is peculiar to transport and refers to the decreasing short-term unit costs which result from the better utilisation of existing indivisible capacity owing to increases in traffic density. Although economies of density are encountered in all forms of transport to a greater or lesser extent, they are of particular importance in the case of rail transport, which usually has large amounts of unutilised capacity. **Economies of scope** are more general and are applicable to multiproduct enterprises. An enterprise of a specific size can claim to have achieved economies of scope if the production of a number of different products by that single enterprise costs less than production by a number of specialised enterprises. The source of economies of scope comprises sharable inputs, such as parts of productive capacity and indivisible equipment which can be used for producing more than one product, production factors which are only partly exhausted by producing one product and inputs which result in inevitable by-products (joint production).

An example of the latter is once again the forward and return journeys making up the return journey. An input can thus be regarded as sharable in respect of the production of a number of products if the products' joint production requires less of the particular input than when they are produced individually, and if the use of other inputs does not increase. As far as transport is concerned, a road haulier can, for example, handle full loads, part loads and perishable freight with his or her existing terminal facilities and

vehicle fleet. If the particular road haulier's total joint cost of handling the three types of freight is less than the total cost would be if three individual hauliers were each to handle the same quantity of each type of freight, then economies of scope can be said to have materialised. The fact that economies of scope occur means that a multiproduct enterprise will be more efficient than a single-product enterprise.

6.5 COST BEHAVIOUR IN MULTIPRODUCT ENTERPRISES

A transport enterprise is unique in the sense that it does not normally provide a single, homogeneous product or service. In the case of air transport there are, for example, business class services, first class services and economy class services. Furthermore, both passengers and freight are transported. This joint production of more than one service or product creates cost apportionment problems for transport enterprises, and makes costing and pricing extremely difficult. As stated above, there is a distinction between joint and common costs. The former arise where a technical association exists, and the latter where no such association exists. In the first case (joint costs) there is usually a fixed ratio; a sheep yields 20 kilograms of meat and 2 bales of wool. In the case of transport, each forward journey undertaken by a vehicle involves a return journey. As regards common production and costs, common facilities are used to produce two or more services without there being any fixed output ratio. The outputs of the services can vary from one case to another and from one period to another. This complicates both costing and pricing in the transport industry.

In the case of joint costs, one can view the forward journey during which freight can be transported as constituting the main product. However, the return journey is a technical necessity and is regarded as a by-product. This means that if we consider the marginal cost of the marginal unit, such cost constitutes the additional, total joint cost (forward and return journey) less the additional revenue which can be earned from the sale of the by-product. For this reason it will always benefit a haulier to endeavour to transport freight during the return journey. This will thus reduce joint costs, because additional income is generated which reduces such costs.

For a transport enterprise faced with joint production, it is thus extremely important to have both the necessary cost and price information. As to how the costs must be divided between the main product and the by-product can really be determined only after the event. Thus, in the case of both of the products or services, the costs concerned can be determined only after the returns of the products or services are known.

As far as common costs are concerned, it is important for the transport enterprise to determine whether it is worthwhile to offer the common services. This means that the common costs of offering the two products together must be lower than where the two products are offered individually by two separate transport enterprises. Think of the road carrier who, for example, offers long-haul and delivery services, or the rail transport enterprise which uses the same track for freight and passenger services. If it would be cheaper to offer one of the services on its own, common production will not contribute to the profit made but will instead reduce profits.

6.6 USER COSTS

There are numerous factors which influence the decisions of passengers and freight shippers. As already mentioned, such persons are not interested merely in financial costs, but also take quality aspects such as speed, reliability and frequency into account. The total opportunity cost of transport is therefore relevant. Although this phenomenon is not unique to transport, transport differs from other services in that price may constitute only a small portion of the total costs, and may play an extremely small part in decision making. This probably explains why many travellers prefer private motor vehicles while they could in fact travel by bus at a lower cost. In analysing or forecasting transport demand it is sometimes possible to evaluate user reaction or individual cost items, but, in most cases, it is useful to make use of a compound criterion. In order to summarise the large number of variables – some of which are not always directly measurable – in a form which focuses attention on trends in total user costs, a compound measure, namely generalised cost, is used. The generalised cost of a journey is usually expressed as a single, linear compound index of individual, dissimilar costs in monetary terms. A characteristic of generalised cost is that it thus reduces all user costs to a single index which can then be used like an ordinary monetary cost in economic analyses. Generalised user costs usually also deviate from the actual source costs of transport. Transport economists are interested in the costs which influence the behaviour of participants in the market in both the short and long term. In the short term, it is in fact possible that transport users will not observe the full range of opportunity costs of their transport use as costs. Nevertheless, it is the perceived costs which influence their immediate behaviour. Users normally underestimate their costs for the following reasons:

(1) Time and monetary costs may be so little that it is not worthwhile taking them into account.
(2) Certain variable costs may be wrongly regarded as fixed costs: Motor vehicle users tend, for example, to take into account only their fuel costs and ignore use-related depreciation and maintenance costs.
(3) Users may be unaware of the relationship between a specific action and the cost brought about by such action: A driver who drives fast may, for example, be unaware that, by doing so, his or her fuel costs increase.

(4) Habit can cause regular users to lose track of changing costs over time, even if they were initially aware of the full economic costs of their actions. This is probably more of a problem in the case of motor vehicle users than in the case of those making use of public transport and who are confronted by increases in travel fares.

Although the concept of a single index which represents transport costs has made an important contribution to the better analysis of transport demand, there are three general criticisms which are levelled at such an approach.

(1) The combination of all costs makes it impossible to determine the demand elasticity of specific items, for example savings in time. Time costs may constitute only a small portion of generalised costs, but may nevertheless exert a considerable influence on the behaviour of users.
(2) The long-term stability of a monetary measure is doubted. Since income increases over time, the utility of money declines relative to other items, such as time, of which there is only a fixed amount available. It is therefore argued that time should rather serve as the measure in the index. Among other things, time is proportionately distributed – each person has only 24 hours in each day at his or her disposal – and such a measure avoids some of the problems related to a disproportionate income distribution. However, the implications of this proposal are not yet entirely clear.
(3) Objections are raised to the use of a general index for an entire country. Although official evaluations of time value and the use of officially approved formulas promote uniformity, these can give rise to inappropriate results if they are valid only in certain circumstances. It is therefore preferable to compile separate indices for different areas.

Despite the criticisms levelled at the concept of generalised cost, such concept remains a useful instrument for understanding the influence of changes in user costs on transport behaviour in general. Although the concept must be used with circumspection, it provides a method of predicting the possible effects of different transport policies (Button 1982:97–102).

6.7 CONCLUSION

The structure, scope and behaviour of costs are determinative for many other aspects of transport. In particular, they are relevant to the economic structure of transport and to the determination of transport prices. You must therefore ensure that you have mastered this study unit before proceeding to study the rest of this study guide.

SELF EVALUATION

(1) Discuss in full the nature of economic costs. (10)
(2) What is the relationship between variable costs and avoidable costs? Discuss. (5)
(3) Discuss in full the difference between fixed costs and variable costs. (5)
(4) Distinguish in full between common costs and joint costs. (6)
(5) Is it possible to apportion common costs rationally to the various outputs that are produced with it? (3)
(6) Why are common and joint costs important to a transport enterprise? (4)
(7) Discuss in full the operation of the law of diminishing marginal returns. Explain your answer with the aid of a graphic representation. (15)
(8) Discuss economies of scale in full. (6)
(9) What is the relationship between large-scale production and economies of scale? Discuss. (5)
(10) Define and discuss economies of density and economies of scope. How can these concepts be applied to transport? (12)
(11) Is transport subject to decreasing average costs in the long term? Discuss. (5)
(12) Are generalised and observed costs suitable measures for analysing user behaviour? (6)

STUDY UNIT 7

The economic structure of transport

CONTENTS

OBJECTIVES

After studying this study unit you should be able to:

- discuss the difference between coordination through direct control and coordination through the market mechanism
- discuss the factors that determine enterprise size
- define a transport market
- discuss the different theoretical forms of competition
- use the cost structure of the different modes of transport to explain the nature of competition within the different modal markets

In the previous study unit you were introduced to the various kinds of costs that are of importance in Transport Economics. These different kinds of costs are responsible for the fact that different kinds of enterprises in the market show different cost behaviour and have different cost structures. The same applies in the transport market. The kinds of costs which enterprises encounter determine whether or not there is competition in the transport market. We shall now proceed to examine this economic structure in the transport market.

The aim of this study unit is to give you an insight into the organisation of transport in a particular market structure and into the factors that affect organisation. We specifically discuss the influence of cost structure and cost behaviour on the market structure and consequently on competition.

7.1 INTRODUCTION

In the previous study unit we emphasised the optimal proportion of fixed to variable factors of production and the optimal absolute size of individual plants. The structuring of factors of production to achieve the optimal ratios and quantities is by no means a simple matter, nor does it happen automatically. It is usually brought about by the market mechanism or by direct control. By whichever

means it is done, however, the organisation of factors or means of production for one purpose or another is the result of human actions, and can be either well-directed or misdirected. There are no hard and fast rules for achieving the best possible organisation of factors of production under all circumstances. In this study unit we deal with a number of aspects of transport economics, of varying importance. Technical matters will not be discussed in-depth, however; we are assuming that transport entrepreneurs are making optimum use of the appropriate technology and processes. If the market mechanism is to allocate scarce factors of production in an optimal manner, there has to be effective competition in the market. The effectiveness of competition is largely dependent on the size of the enterprise in relation to the extent of the market in which it is operating. For this reason we begin by discussing enterprise size, followed by a discussion of the characteristics of transport markets, and, lastly, transport competition.

7.2 SIZE OF THE ENTERPRISE

7.2.1 Introduction

If the law of diminishing returns did not operate, and there was no limit to the economies of scale that could be achieved in all production or operating processes, in other words if unit costs continued to decrease in both the short and the long term, economic organisation would be entirely problem-free. Under such conditions, the optimum utilisation of factors of production would involve the production of the particular product or service by a single enterprise, because unit costs could be continually reduced by the expansion of production. The structure and behaviour of various transport services means, however, that enterprises of various sizes, at different stages of diminishing returns and returns to scale, compete with one another in a transport market that does not necessarily function efficiently.

7.2.2 Methods of coordination

The transport industry in a country can be seen as consisting of a large number of operating processes, such as the operation of a bus on a particular traffic route, the signal control system of a train, the building of a locomotive or the assembly of a truck. The individual processes involved in providing transport services are really too numerous and varied for us to give a comprehensive account of them. The question that naturally arises is how the process should be planned and organised. Each process contributes to the provision of a transport service and should be carried out in conjunction with all the other processes that have the same purpose.

From the point of view of transport economics, the aim is to provide the required transport services at the lowest possible economic cost. Two possible approaches can be distinguished. First, each process could be separately controlled or managed and the coordination of the different but related processes required in order to provide the particular transport service could be left to the forces of supply and demand, which operate through the price mechanism. Alternatively, the various related processes could be placed under the direct control of a single entity as a means of effecting coordination. If the operation, garaging and repair of a bus are seen as three separate processes, it is clear that each process could be separately controlled, in which case the bus operator would rent space from an independent garage and would pay to have his/her bus repaired by an independently operated workshop. All three of these processes could, however, be combined under the direct control of a single enterprise.

For the purposes of this discussion an enterprise could be regarded as a group of processes which are under the control of a single decision maker. The decision maker could be a single entrepreneur or the supreme decision maker in the enterprise.

In view of the large number of processes associated with transport, it is clear that numerous combinations of processes could be found in enterprises. One enterprise could assemble buses and another could operate them, or a single enterprise could do both. Similarly, one enterprise could undertake the line-haul transport of consolidated parcels and another a parcel delivery service or both processes could be undertaken by a single enterprise. A rail transportation enterprise could buy its locomotives from an enterprise that manufactures mechanical equipment or else build them itself. We can conclude that the size of the enterprises depends on the way in which different processes are grouped into separate enterprises or grouped under separate control. If we accept that a fixed number of processes are involved in satisfying the total demand for transport, it follows that the larger the number of processes that are controlled by a single entrepreneur or enterprise, the smaller the number of enterprises that will be in existence but the bigger those enterprises will be. But it is naturally also true that the smaller the number of processes that are grouped together, the larger the number of small enterprises there will be. Whatever the particular groupings may be, the combined processes within an enterprise are coordinated by means of direct control, and the coordination between different enterprises in a market economy is accomplished through the price mechanism. If there are a few big enterprises supplying the total demand, coordination takes place chiefly through direct control, but if there are a large number of small enterprises supplying the market demand, the price mechanism is the most important method of coordination.

From the point of view of transport economics, any study or assessment of the size of transport enterprises should be based on the relative economic merits of direct control as opposed to the price mechanism. The size and number of enterprises capable of supplying the aggregate demand for transport at the lowest economic cost should be taken as the criterion. Because competition between a large number of small enterprises is regarded as the ideal in a free market system, it is important that the advantages and disadvantages of larger as opposed to smaller enterprises should be more closely examined.

7.2.3 The role of economies of scale

It is undoubtedly possible for larger enterprises to use scarce factors of production more efficiently than smaller enterprises. You have already come across the explanation for this, namely the existence of economies of scale and economies of scope. In the first case the economies are achieved by using plant and equipment with a greater capacity. Although a bigger plant or more extensive equipment are probably more expensive to acquire and install, their production capacity can be so much greater than that of a smaller plant that this more than compensates for the higher costs. There is an accompanying drop in unit costs. A double-decker bus is cheaper to run than two single-decker buses with the same capacity. If a single-decker bus is too small whereas a double-decker bus is big enough to meet the demand on a particular route, it is more cost-effective for one enterprise to operate a double-decker bus on that route than it is for two different enterprises to operate two single-decker buses, because the two smaller enterprises would be unable to take advantage of the potential economies of scale. Similarly the unit cost of a single garage capable of containing twenty buses is lower than that of twenty garages that each has the capacity for only one bus. Furthermore, the unit cost of bigger, specialised workshops to repair buses is probably lower than that of a number of smaller workshops. As you already know, however, no plant remains permanently in a stage of increasing returns, and there are distinct economic and technical limits to economies of scale. The market may be too small to absorb the optimum output of a big specialised plant. A parking garage that is capable of housing fifty vehicles is probably more economic than one that can accommodate ten vehicles, but it offers no economic advantage if as few as ten vehicles are required to meet the total demand for transport in the particular market. The volume of demand or size of the market places a limitation on the size of the plant that can be advantageously utilised. There are a further two kinds of technical limitations on economies of scale as regards plant size.

(1) The technology in question can be subject to inherent constraints. The operating costs of vehicles, aircraft and ships normally increase rapidly after a particular carrying capacity has been reached, and further growth can in fact result in diseconomies.
(2) Economies of scale can be technically restricted by the capacities of complementary equipment and plants that are used together with the plant in question. The volumetric capacity of trucks, that is their cargo space, may be restricted by bridge clearances, and tonnage may be restricted by the structural strength of road surfaces and bridges. Similarly, the limitations on permissible axle mass loads prohibit the use of bigger goods trucks in rail transport. The economic and technical limitations on economies of scale operate alternatively. If the economic limit applies, a single plant can meet the entire demand, but if there is a technical limit, more than one plant has to be commissioned to supply the aggregate demand. If twenty passengers on a certain route make use of bus transport at a particular time, a single-decker bus can be used and the use of a double-decker holds no economic advantages. The economic limit, namely the market demand, applies here because it is lower than the technical limit, namely the capacity of a double-decker bus. If, however, the demand on the particular route consists of two hundred passengers, the demand exceeds the carrying capacity of a single-decker bus, and a number of double-decker buses may have to be used. In this case the technical limit predominates because the plant in question cannot be sufficiently expanded to meet the total demand and more than one plant of optimal size would have to be used.

7.2.4 The role of economies of scope

If the individual plant or process is used as the criterion, there is no reason why any enterprise should be larger than is necessary to control a single plant of optimal size, which would be a plant that, under prevailing circumstances, offers no more potential economies of scale. If the economic and technical limits in respect of scale are high, the enterprise in question will be correspondingly large. In the case of electricity generation and steel manufacturing the economies of scale are sufficient in themselves to justify enterprises with a large capacity. When it comes to transport, with the possible exception of rail transport, the limits on the scale of plants are such that the potential economies of scale are soon exhausted, even by a relatively small enterprise. A transport enterprise with a single plant is one with only a single vehicle, ship or aircraft. Although an enterprise that operates a ship is larger than one that operates a truck or an aircraft, they are all nevertheless relatively small. If economies of scale as a result of the use of larger plants were the only advantage of large enterprises, and the size of the enterprise consequently depended on the nature of the plants employed, the size of enterprises would have varied from very small to small, and not from very small to very large. There are therefore other factors that make larger enterprises more advantageous and more competitive.

The additional factor that enables larger enterprises to make better use of their factors of production is the better coordination between different plants or processes. If two or more processes produce a particular output in close association with one another, the processes can often be more efficiently organised if they are controlled jointly by a single enterprise. If decisions with regard

to one process have consequences for another process, a single decision maker may be able to achieve the desired results at a lower cost. The single decision maker might be able to spot potential economies of scale that might not be obvious if the different processes were under the control of separate enterprises. If the output of the different processes can be regarded as the products of a single enterprise, even if some of them serve as inputs for the operating or production process, we are in fact dealing with a multiproduct enterprise. The economies which make it advantageous to group a number of related processes within a single enterprise are therefore no different from those described earlier as economies of scope.

In the case of railways the maintenance of the railway track, the control of the use of the track and the operation of trains are indeed outputs that could just as easily have been produced by different enterprises. They are, however, so closely related that their coordination by a single enterprise is economically advantageous. The primary saving would probably be on management costs, but other costs would not be higher than when separate enterprises were controlling the processes. In the case of roads, however, the maintenance of roads and traffic control are not so closely related that there is any benefit to be derived from their joint control. The situation is different when it comes to road transport operators, however. Factors such as the transfer of vehicles from one route to another when peak demand times on the various routes do not coincide or the smaller number of reserve vehicles needed to cover breakdowns and the servicing of vehicles or to cater for peak demand make it economically more efficient to run bigger networks. The consolidation activities of road transport enterprises that cater for less-than-truckloads or parcels represent a by-product that can be advantageously combined with other transport operations. Direct control can also promote the physical coordination of transport, for instance by synchronising the timetables for different services.

When the grouping of different services under the direct control of a single entity is being considered, the question of vertical and horizontal integration arises. Horizontal integration takes place when similar processes such as long-line transport over different routes are placed under the control of the same entity in order to form a larger "long-line" enterprise. But if the processes in question are dissimilar and form successive links in a chain of processes, vertical integration takes place. An example here would be where a manufacturing or mining enterprise undertakes its own transport. The coordination of transport processes by means of direct control, usually by an authority or government organisation, constitutes horizontal integration. As in the case of economies of scale, the question is whether there are any limits on the economies that can be achieved by combining associated but dissimilar processes under direct control, in other words whether economies of scope are possible. If there were no limits, larger enterprises would always have been economically more efficient than smaller enterprises. However, any enterprise reaches a point where diseconomies of scale set in. Although management can be decentralised, the scope and variety of decision making begin to affect the ultimate decision making entity at one time or another, irrespective of whether a single entrepreneur or a whole board of directors is concerned. The rate of decision making regarding individual processes slows down and it becomes impossible to apply a consistent policy. The optimum size of an enterprise is ultimately determined by the balance between economies of scale and economies of scope on the one hand and diseconomies of scale on the other. It is impossible to lay down general rules but the unit costs of an enterprise are a reliable measure of whether it has attained its optimum size.

7.3 ATTRIBUTES OF THE TRANSPORT MARKET

7.3.1 General

In economic theory a market is defined as any place where buyers and sellers meet at any time to buy and sell goods and services. Economists studying the behaviour and performance of an economic system generally concentrate on the characteristics of individual enterprises, since it is through the behaviour of these enterprises that the general objectives of the particular economic system are achieved. However, where the emphasis is on individual enterprises, it is possible to lose sight of the overall economic picture. The development of industrial organisation as a field of study is an attempt to find a golden mean by studying enterprises with reference to their interaction in the market. This interaction chiefly takes the form of market competition. From the point of view of industrial organisation, the study of competition in the market centres around the structure of the market and the behaviour and performance of enterprises in the market.

Market structure refers to the economically significant attributes of the market which influence the behaviour of sellers. It embraces factors such as the number and sizes of buying and selling enterprises (as discussed in section 7.2) in the market, constraints on market entry, the role of fixed costs and constraints on withdrawal from the market, and the degree of product differentiation in the market.

Market behaviour refers to the behaviour of an enterprise in relation to its product market and its competitors. Some of the most important elements of market behaviour are price policy, the evaluation of product or service quality and the discouragement of market entry or the exercise of pressure on existing competitors.

Market or economic performance is usually measured by the degree to which an enterprise contributes to the basic objectives of the economic system, such as the efficient use of scarce factors of production, progress, economic growth and an equitable distribution of income. In the discussion that follows the emphasis falls chiefly on market structure.

7.3.2 Definition of the term "transport market"

The definition of a market structure is very closely related to the size of the market, and we therefore need to define the term "transport market" before examining the structure of transport markets. We could take the particular product or service as the point of departure. It must be possible to substitute the products of one enterprise for those of another before the two enterprises can be said to be competitors for one another. The perimeters of a market are therefore determined by the homogeneity or substitutability of the products traded in that market.

It was formerly believed that transport markets could be divided into two separate markets: One for passenger transport and one for freight. It was also assumed that each of these markets produced a homogeneous product, namely passenger-kilometres and ton-kilometres respectively. It has now become customary to refine this classification on the basis of three variables, namely geography, the kind of service or product produced and the quality of the service.

(1) Transport markets can be distinguished on the basis of the geographic areas or points they serve, in particular the relevant origin-and-destination pairs served. A transport service between Johannesburg and Durban is not the same as one between Johannesburg and Cape Town, to take one example. Such services can seldom be substituted for one another, except possibly in the case of circular routes. Services over routes that partly coincide can be partial substitutes for one another, provided they connect with other services. Outward and return legs of the same route are not substitutes and in fact represent two different geographical markets.
(2) Transport markets can be distinguished from one another on the basis of the kind of transport service offered. In cargo transport, one can distinguish between the transport of full loads (consisting of loose or packed bulk cargo), the transport of less-than-truckloads or full loads of general merchandise, with the first kind predominating, and the transport of less-than-truckloads consisting of parcels. In the case of passenger transport it is impossible to make a clear-cut distinction between the two kinds of service, because the distinction tends to depend on service quality. The most important distinction would probably be that between individual and bulk transport.
(3) Service quality or service level is also a criterion for the identification of separate transport markets. (The quality of transport services is discussed in greater detail in the study guide for TRL201X.) Transport services can differ considerably in respect of quality requirements relating to speed, reliability, punctuality, safety and the extent to which operators accept financial liability for the damage or losses incurred by users. Some of these requirements create an opportunity for specialisation and product differentiation, such as where express transport services are offered. It is apparent from the above that transport enterprises offer a wide variety of services, and that individual enterprises frequently provide more than one kind of service in the market. When transport markets are described in terms of all three of the relevant variables, it becomes apparent that individual markets are probably a lot smaller than is generally believed. As regards market structure, it means that the enterprises of individual operators may be large in relation to the markets they serve, and consequently a market may be limited to only a few suppliers. The implications of these conclusions will become clearer in the following sections, which deal with market structure and competition.

7.4 FORMS OF COMPETITION

7.4.1 General

In theory "competition" can take many different forms, and the degrees of competition range from perfect competition to pure monopolies. Viewed superficially, competition consists of rivalry between the suppliers of goods and services in order to maximise their income or profits. The attainment of this goal is restricted by the extent to which buyers are able to obtain substitutes from alternative sources on more attractive conditions. From an economic point of view it is important to establish in which circumstances competition would lead to the optimal coordination of scarce factors of production. The most important forms of competition will be briefly discussed in the following sections:

7.4.2 Perfect competition

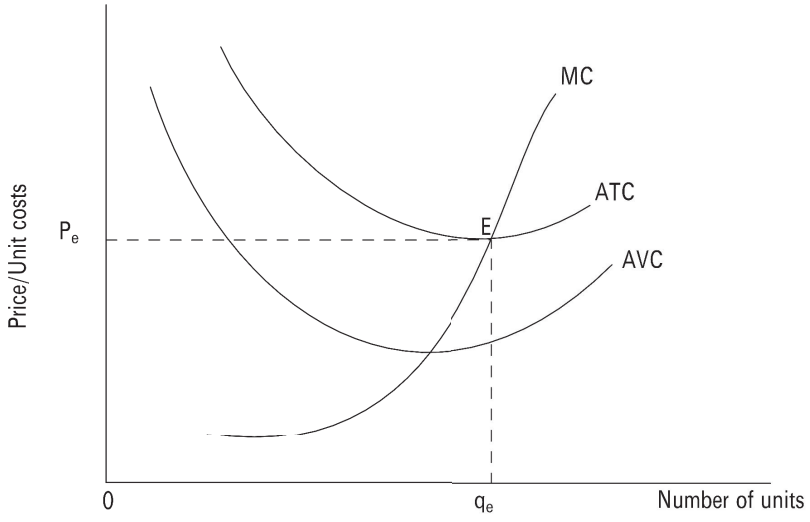
Perfect competition is a theoretical concept and is not encountered in practice. In a perfectly competitive market there are no obstacles to the optimal or efficient utilisation of scarce factors of production. In a market of this kind, firstly, only homogeneous

products or services that are completely substitutable are sold. Secondly, there is a sufficient number of buyers and sellers, so that no single buyer or seller can influence the market to his own advantage, and there is no collusion to attempt to exert such influence. And, thirdly, no constraints are imposed on the entry or exit of buyers and sellers. The last requirement implies that the factors of production are perfectly mobile and can be freely transferred to other applications or markets where they could earn higher returns.

The above requirements have important implications for the operation of the market. Because individual suppliers are small in relation to the market, they are able to sell any quantity of their product or service at the prevailing market price, which means that they are confronted by a perfectly elastic (horizontal) demand curve. Furthermore, the individual sellers must accept the price set by supply and demand in the market as given, and adapt their production policy and costs accordingly. In other words they are price takers. Under conditions of perfect competition cost and price are therefore equivalent to one another at the equilibrium position, as illustrated in figure 7.1. A supplier will expand his production as long as his marginal cost is lower than the ruling price, because he can earn additional profits in this way. Similarly, he will reduce his production if his marginal cost is higher than the market price since the production of an additional unit would result in a loss. The price would therefore be equivalent to the marginal cost. The pressure of competition also means that individual suppliers of products or services produce at the lowest possible unit cost. At this point the unit cost or average total cost is equal to the marginal cost. The reason is that when marginal costs are lower than average costs, the average costs decline as output increases. However, marginal costs begin to increase when the stage of diminishing returns is reached. A point is reached where average cost and marginal cost are equal, after which marginal cost begins to exceed average cost, with the result that the average cost also begins to rise. Average cost therefore reaches a minimum when it is equivalent to marginal cost. No enterprise can continue indefinitely to produce at an average cost that is higher than the market price, because this would result in a loss and the enterprise would then be obliged to leave the market. Conversely, no enterprise can indefinitely charge a price which is higher than its average cost, because the surplus profits made would cause the supply in the market to rise because existing suppliers would increase their offering and new suppliers would enter the market. The higher supply would cause the price to drop until it is equal to the average cost. Short-term equilibrium, point E in figure 7.1, is therefore achieved when price, marginal cost and average cost are equal to one another. Although they are not discussed here, the same conclusions apply to long-term equilibrium.

FIGURE 7.1

Short-term competitive equilibrium



7.4.3 Pure monopoly

On the opposite side of the spectrum of forms of competition, we find pure monopoly. This involves a market structure consisting of only one supplier of a product or service for which there are no substitutes. The most important implication is that the monopolist is no longer a price taker, but is able to influence the price directly by increasing or reducing his output. If he offers more of his product or service on the market, the price of his full offering drops, and if he reduces his offering the price of his full offering

increases. If a monopolist is considering expanding his output, therefore, he must consider the relationship between the marginal cost of his additional production and the price, that is to say the marginal income, he could receive for it. If his marginal costs are lower than the expected marginal income, he could increase his profit by offering more of his product or service on the market. But because the monopolist has to reduce his price in order to sell his additional output, he will not be able to go on increasing his profits indefinitely. It appears that the monopolist will have to manipulate his supply and his price until his marginal costs and his marginal income are equal to each other.

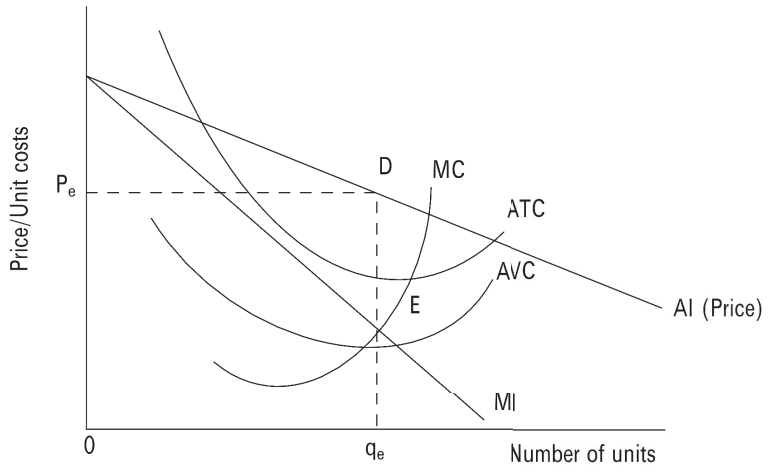
Because the market price declines in proportion to any increase in the monopolist's income, the marginal income curve (MI) invariably lies below the average income curve (AI). As is evident from figure 5.2, this means that the points at which the marginal cost and marginal income curves intersect always lie below the average income curve. An important consequence of this situation is that the price will always exceed the marginal cost at the point of equilibrium (DE in figure 5.2). This does not mean that the monopolist will always make a profit, since it is possible that the average income curve (AI) could lie below the average total costs curve (ATC).

A pure monopolistic market structure is almost as rare as perfect competition. It is more practical to investigate the market structure in order to establish the degree of monopolistic power exercised in it. Monopolistic power enables a seller to influence the market to his own advantage by means such as the manipulation of output and price. The degree of monopolistic power possessed by an enterprise may not (as became apparent above) be determined without reference to the buyers of his product or service. The price elasticity of the demand for his product determines the extent to which the monopolist is able to use his power to influence the market to his own advantage. A highly elastic demand largely limits the possibility of making a monopoly profit.

The above discussion applies mainly to short-term monopolistic behaviour. Because there is no question of market entry, the analysis of long-term monopolistic behaviour is conducted in a very similar way. The most important differences between the two probably lie in the long- and short-term cost functions.

FIGURE 7.2

Short-term monopolistic equilibrium



7.4.4 Natural monopoly

Traditionally the inherent tendency towards declining unit costs over the entire demand range of the market is regarded as the distinctive characteristic of a natural monopoly. In a case like this ensuring the optimal use of the factors of production would simply require one enterprise to serve the entire market. The indivisibility of some fixed factors of production is regarded as the most important source of diminishing unit costs. In some industries a substantial investment is required before the demand of any one buyer can be satisfied. For instance, a railway enterprise would have to build a railway line between two points before any service could be supplied. The costs involved are fixed, irrespective of the size of the eventual output, and the cost per unit consequently decreases as the output increases. The arguments in favour of having a single enterprise serve the market where indivisibility is a factor are supported by some associated attributes of monopolistic enterprises: The indivisible factors of production or facilities are usually geographically immobile and cannot be moved once they have been created; the facilities are likely to be specialised

and the associated costs are largely irrecoverable, in other words sunk; the services (and in some cases the products) concerned cannot be stockpiled; the service has to be supplied on demand, and the demand is subject to the peak phenomenon. The output capacity of the facilities must therefore be sufficient to meet the peak demand when it arises, and it is therefore economically more efficient to have a single enterprise supplying the market.

A single seller can also represent the optimal market structure in cases where technical economies of scale apply. The only requirement is that the larger plants must have a lower unit cost than smaller plants or that the expansion of existing capacity should cost less than the creation of new capacity – take the example of doubling the size of an existing railway line rather than building a new line. The investment involved need not necessarily be extensive or fixed; instead it could be proportional to the growth in demand.

The tendency towards diminishing unit costs is clearly related to the fact that fixed or capacity costs represent a very large proportion of total costs. A high proportionate fixed cost is, however, not sufficient to ensure that an enterprise will enjoy a natural monopoly. If the unit costs of an enterprise do not decrease over the full demand range as a result of indivisibilities or technical economies of scale, production could just as well be undertaken by a number of smaller enterprises. The existence (and protection) of natural monopolies is therefore related to the waste that would result from competition if certain facilities were to be duplicated. It is certainly possible that a market could be served efficiently by a number of enterprises if the extent of the demand is large enough for optimum use to be made of more than one indivisible facility or for the technical economies of scale of the individual enterprises to be exhausted. Duplication is only undesirable when the aggregate demand is such that factors of production are optimally utilised when the market is served by a single enterprise.

7.4.5 Other forms of competition

Between perfect competition and pure monopoly we find a number of other forms of competition; they are defined in terms of the specific requirements of perfect competition from which they deviate and the extent to which they deviate. For example, oligopolistic competition is said to prevail where there are only a small number of sellers in the market. In a market of this kind sellers influence one another's decisions through their actions. The marginal income a seller can obtain by boosting his output is dependent on the future reaction of other sellers – an unknown factor – and the determination of his marginal costs is dependent on his assumptions regarding the possible reaction of competitors. An oligopolist therefore finds it more difficult than a monopolist to manipulate his price, and in fact it is impossible to develop a general theory of oligopolistic competition. If oligopolies form an effective cartel in collusion, they are able to manipulate the market in the same way as a monopolist. On the other hand, if they function completely independently, the market could function as a competitive market.

Another fairly common form of competition is monopolistic competition. It differs from perfect competition in that it does not meet the requirement of a homogeneous product or service, and it involves competition between a large number of sellers of similar, but differentiated, products or services. In a market of this nature we distinguish between products or services that are essentially substitutes for one another, but differ slightly in respect of factors like trademark, packaging and sales conditions. Each enterprise has a degree of monopolistic power in respect of its own product, and this enables it to make a certain monopolistic profit.

7.5 WORKABLE COMPETITION

Because of the strict requirements for perfect competition, even the most effective forms of competition found in practice could be described as imperfect. It is therefore useful to identify the forms of competition that are found in practice and that do lend themselves – under certain conditions – to the optimal application of scarce factors of production. This requires a study of the types and degrees of imperfect competition which nevertheless produce satisfactory results. Competition that meets this requirement is known as workable competition.

Workable competition can be explained as follows: Given the existence of a number of essential requirements for perfect competition, it does not necessarily follow that the adverse results of a departure from one of these requirements can be reduced by the presence of the other requirements (Clark 1940:250). Indeed, the full presence of the remaining requirements could even reinforce the impact, whereas some degree of imperfection with regard to one or more of these requirements could help to produce satisfactory results. For example, perfect competition requires production to take place at full capacity, where full capacity is defined as the point where minimum unit costs are minimal or the point where average total costs and marginal costs are equivalent. In reality enterprises that are subject to demand fluctuations frequently produce below full capacity for long periods. Because of the existence of unutilised capacity, the marginal or incremental cost of additional production is considerably lower than the average cost. As explained in section 7.4.2, with perfect competition the price of a product simply covers its marginal cost, which means that enterprises that produce below full capacity for extended periods will probably never cover their full capacity. In such cases the perfectly elastic or horizontal demand curve of perfect competition is not reconcilable with workable competition. What is required is a negatively sloping demand curve with a steep enough slope to ensure that the price is high enough above average

total costs to cover average costs during periods of both high and low demand. A further requirement is sufficient price elasticity to stimulate demand in bad times and discourage it in good times. This means that the negatively sloping demand curve which is associated with monopolistic power, and which is the product of market imperfections such as monopolistic or oligopolistic competition, would in this case help to make competition workable. Workable competition therefore does not need to comply with all the requirements for perfect competition.

The definition of theoretical forms of competition emphasises price and output to the exclusion of dynamic competition, which also covers competition in respect of aspects such as product or service design, production costs and sales promotion. Potential competition by new entrants and the substitution of products or services also help to make competition workable under imperfect market conditions. The behaviour of potential competitors who may or may not already be active in the market can have an inhibitory effect on monopolistic profit making. Such behaviour can take the form of entry to the market, the introduction of new products or services, the introduction of more efficient production or operating methods, new sales promotion techniques or price reductions. Workable competition does not require a large number of sellers – two or more entrepreneurs functioning independently of one another are sufficient. Even the development of a monopoly or strong cartel does not necessarily mean the end of competition because the threat of competition by potential entrants still exists. The development of substitutes means that the demand curve for individual enterprises with monopolistic power becomes more elastic, which counteracts high monopoly prices or price discrimination. The development of road transport as a substitute for rail transport in some markets substantially increased the price elasticity of the demand for rail transport in those markets. The concept of dynamic competition helps to explain the workings of competition under actual conditions.

7.6 CONTESTABLE MARKETS

The theory of contestable markets is a recent development that has attracted a great deal of attention in Transport Economics. You are referred to authors such as Taplin (1983), Davies (1986) and Tye (1987). This theory formalises earlier intuitive views on the role of potential competition in helping imperfect markets to function adequately. There are fewer requirements for perfect contestability than for perfect competition, but they are equally stringent and are even considered unrealistic by some authors, such as Shepherd (1984). However, this does not in any way detract from the usefulness of the theory and the insights which it has provided into the contestability of the markets.

The most important requirement for contestability is what is known as costless reversible entry. One of the basic premises of this theory is the assumption that it is sunken costs – and not economies of scale – that restrict entry to the market and give the market occupants monopoly power. The risk attached to the investment of large sums (which cannot be recovered upon leaving the market) in specialised facilities deters new investors from exploiting what might otherwise have been profitable opportunities. The effectiveness of potential competition as a disciplinary force increases in proportion to any decrease in irrecoverable entry costs. If a market is easily contestable, competition for the market is just as effective as competition within the market in preventing the exercise of monopoly power, even if there is only one seller in the market.

It seems, therefore, that contestability promotes the optimal utilisation of factors of production across a wide spectrum of market structures. An additional requirement applies, however, namely that the market structure should be sustainable. Sustainability requires the existence of a price, for single-product enterprises, or a set of prices, for multiproduct enterprises, at which

(1) the total market demand can be satisfied
(2) all the enterprises that are active in the market can cover their production costs
(3) prospective contestants are denied any opportunity for profitable entry to the market

In the case of a multiproduct enterprise, sustainability should make provision for the different methods of entry. A new entrant might specialise in one or a small subgroup of the market occupants' products. If there is cross-elasticity of demand between the products, and especially if they are substitutes, the new entrants could even supply larger quantities than the market occupant to one or more of the markets. For example, a new bus operator might offer a larger capacity on a particular route than a multimodal operator who operates buses, trains and aircraft, and he might attract passengers away from all three of the multimodal operator's modes. This can only happen, however, if the multimodal operator's prices for his various services fail to ensure the sustainability of his market.

The theory of contestable markets probably requires a substantial amount of research before it can gain general acceptance but it does nevertheless provide useful guidelines for the development of a competition policy for transport. In Oum et al (1995: 483–499), Harrison and Winston elaborates on contestability. They distinguish between **perfect contestability** and **imperfect contestability.** The first relates to costless reversible entry and very little or no sunk costs. The second **imperfect contestability** is where there might be costs and specific sunk costs, but imperfect contestable markets can still improve economic welfare. They investigated the domestic airline market in the USA and came to the conclusion that it is not a perfect contestable market, but an imperfect contestable market.

7.7 COMPETITION IN THE TRANSPORT MARKET

7.7.1 General

Since the introduction of mechanised transport in the previous century, structural defects in the transport market have been one of the main reasons for the regulation of transport. It has long been accepted that rail enterprises tend to be natural monopolies, whereas oligopolistic market structures occur in rail transport, sea transport and specialised forms of road transport. These structural defects give rise to undesirable behaviour on the part of the enterprises, such as price discrimination, non-price competition, prevention of new entries to the market, the pursuit of objectives other than long-term profit maximisation, and ruinous competition. The latter is a particularly important factor in transport.

7.7.2 Ruinous competition

When looking at this form of detrimental competition it is important to keep in mind that there are three aspects which are important. Firstly there are certain reasons for ruinous competition to take place. Secondly there are certain negative results that follow when ruinous competition takes place. Lastly there are negative consequences for the market structure when ruinous competition has run its course. In South Africa the intense competition in the mini-bus kombi-taxi industry after deregulation is a case in point. The ruinous competition has forced the government and transport authorities to introduce new regulations as well as the actual taxi recapitalisation programme.

Ruinous competition arises when two or more enterprises with unavoidable excess capacities compete for the same market. This is usually the case with transport enterprises that offer scheduled services and have a relatively high proportion of fixed or sunk costs in relation to total costs. Large fluctuations in demand and a low price elasticity of demand aggravate the situation. As a result of the availability of unutilised capacity the marginal or incremental costs are considerably lower than the average costs because the increase in the output requires only additional variable costs.

Owing to the presence of sunk costs and the immobility of factors of production, capacity cannot be adapted promptly in response to changes in demand and enterprises are compelled to compete in the market in order to utilise their excess capacity. Unrestrained price rivalry forces prices down to the level of marginal or average incremental costs, and when the price elasticity of demand is low the end result is a lower income for all the competitors. In these circumstances price competition often takes the form of price wars, secret rebates and discriminatory prices. Some enterprises may resort to using the monopoly profits they make in other sectors of the market to finance their price cuts in a ruinously competitive market. In relatively unconcentrated markets, where there are too many small undertakings to permit a coordinated attempt to prevent prices from being forced down, there is a possibility that the whole industry, or at least a large part of the industry, will operate at a loss for extended periods. Capacity which may be required in the long term might have to be sacrificed in this process. In an oligopolistic market excessive price competition often results in price agreements that enable the enterprises concerned to recover their losses.

History has shown that in the absence of government intervention industries that are inclined to ruinous competition do not remain depressed indefinitely. They do, however, have a history of unusual instability as far as income goes, in both the long and the short term. The underlying cause is the inelasticity of supply in these industries. A considerable period usually elapses between the taking of a decision to extend capacity, such as the duplication of a railway line or the building of a new ship, and its final implementation. Because supply is slow to react to an increase in demand, prices tend to increase rapidly as the demand increases. There is a serious problem if the anticipated growth has not taken place when the capacity eventually becomes available on the market. The industry would then have unutilised capacity, which would lead to price competition and consequently to lower prices. The same situation would arise if the market demand was to decrease significantly and the capacity could not be reduced promptly. The continued excess capacity that has been plaguing the oil tanker industry for some time is a case in point.

Price competition is not in itself undesirable and is in fact the mechanism by means of which a competitive market gets rid of inefficient excess capacity. In the case of ruinous competition it is detrimental to the user or buyer in the long term. One reason is that reduction in the seller's income may cause him to suspend postponable expenditure. The maintenance, repair and supply of capacity that the market requires in the long term and that will cost more to provide later on, as well as the research and development and the supply of temporarily uneconomic products or services, could be discontinued. This situation is especially likely to occur when market defects are present, such as imperfect foreknowledge of demand levels on the part of producers, and the limited mobility of fixed factors of production. This creates the possibility that, for relatively long periods, the return on capital will be insufficient for the maintenance of capital in the long term.

Major price fluctuations in industries prone to ruinous competition are not in the interests of either sellers or buyers. Such fluctuations hamper long-term planning and promote speculation at the expense of efficient production or operating procedures.

Consumers' limited ability to assess quality and keep it at an acceptable level is a further imperfection of the market and one that is to the detriment of consumers under conditions of ruinous competition. The quality of transport services has several dimensions, as you already know. Apart from the physical aspects, factors such as safety, punctuality, reliability, high frequency and financial liability accountability are also important to the consumer. The decrease in the income of suppliers under conditions of ruinous competition may lead them to neglect these quality requirements, and the greater the inability of users to notice any reductions in quality immediately, the greater the temptation will be for suppliers to economise in this area. Naturally it is possible to protect the interests of the consumer by enforcing quality standards by means of legislation (Kahn 1971:172–178; Sharkey 1982:24–28).

7.7.3 Competition within the various modal markets

Virtually every kind of competition is found in the transport market. Since the type of competition that occurs in a particular mode is dependent on the cost structure of that mode, there are naturally differences between the various modes as regards competition.

In air travel the economic limit on economies of scale usually applies, and most air transport enterprises operate more than one airline. This means that large enterprises will only be able to make optimal use of factors of production if economies of scale with respect to enterprise size are present. Although technical economies of scale are not found in air transport, increases in the size of enterprises could be justified by the presence of economies of scope. It may be more economical for one airline to undertake both scheduled and charter flights than for various specialised air transport enterprises to undertake different types of flights. In the United States of America, where competition in the air transport market is no longer strictly regulated, there appears to be a tendency towards an oligopolistic market structure. In most countries, air transport is still regulated, however, and the market structure is the result of regulatory policy rather than the operation of economic forces.

Market structures in the ocean shipping industry vary from almost perfect competition in the case of tramp shipping to oligopolistic cartels in the case of shipping conferences. It is unlikely, however, that ocean lines that cater for shipping conferences could ever function as monopolies, owing to the competition from tramp shipping. Economies of scale are possible in the case of large vessels, but not necessarily large shipping enterprises. This enables small or one-ship companies to compete with large ocean transport enterprises. In a study of Australian shipping lines, Davies (1986) found that the market largely met the requirements of contestability: Sunk costs are low in general, contractual prices meet the requirements for sustainability, and there are a large number of potential competitors. Although factors such as the existence of national shipping lines and political intervention can impose limits on contestability, the actions of private shipping lines meet the requirements for a contestable market.

As regards rail transport, the most recent studies show that the long-term cost curves of rail transport are horizontal, which implies constant returns to scale. As far as the market structure of rail transport is concerned, it is essential to distinguish between decreasing unit costs due to economies of scale and those due to economies of density. Owing to economies of density a railway enterprise could come to enjoy a natural monopoly in respect of a particular route. If the expansion of capacity through the linking of various routes fails to achieve any economies of scale, however, this means that there is a natural monopoly in existence for each route, without the railway as a whole being a natural monopolist. Most of these arguments are largely academic, however, because most countries in the world have only one railway enterprise.

As in ocean shipping, the market structure of road transport, especially road haulage, ranges from near perfect competition to a highly oligopolistic structure. The market for truckloads of general freight or packaged bulk cargo loads best satisfies the requirements for perfect competition. Fleet sizes in this market vary from one to a hundred vehicles. In the case of single truck operators sunk costs are very low and this market largely complies with the requirements for contestability, although there is a tendency to move away from a competitive structure. Specialised carriers and carriers of less-than-truck loads and parcels generally require costly terminal facilities that raise their fixed costs and lead to declining unit costs with increased usage. There are also other factors that favour larger road transport enterprises. Larger carriers have the advantage in matters such as insurance, and the ability to attract capital and solicit for freight. Freight soliciting plays an important part in carriers' efforts to achieve a balanced traffic flow and in consequence make better use of equipment and facilities.

Economies of scale in road transport are still a debatable issue. If only vehicles are taken into account, road transport is probably subject to constant economies of scale. But when the indivisible assets of specialised carriers – such as sorting facilities and storage facilities – are taken into account, the likelihood of achieving economies of scale increases. Other potential sources of economies of scale for larger road carriers are the use of specialised vehicles, the use of their own workshops for maintenance and repairs, cheaper purchase of inventory, and specialisation, especially in the area of management. However, all these advantages do not preclude competition from small operators.

Lastly, we can take a brief look at pipelines. It is widely accepted that pipelines are subject to economies of scale. If the length of a pipeline is kept constant, any increase in the diameter of the pipe leads to a decrease in the unit cost of pipeline transport. It appears, however, that there are very few savings attached to an increase in the length of the pipeline. Although the presence of economies of scale implies a negatively sloping long-term average cost curve, it does not mean that a large pipe is necessarily

more efficient than a smaller one. If the full benefit is to be derived from the potential economies of scale, a high output over a long period would be required. The most important characteristic of the cost structure of any pipeline transport service is that fixed costs are proportionately much greater than variable costs.

In this section we have simply examined competition within the various modes. Competition naturally also occurs between different modes of transport. We call this type of competition intermodal competition. A possible example would be competition between coastal shipping and rail transport for the cargo trade between Durban and Port Elizabeth. There could also be competition between road transport and rail transport, for instance for the transportation of cargo between Johannesburg and Durban. In this case various modes of transport are competing for the same cargo.

The various modes of transport and their respective economic attributes are fully discussed in the third-year course in Transport Economics, TRL300.

7.8 CONCLUSION

It is impossible to do justice in the scope of a single study unit to the whole question of the economic management of transport and competition in the transport market. We have tried to show the most important aspects in order to give you an overall picture of the subject, because this relates to the relationship between the costs of an industry and its pricing strategy. Price formation in transport is the topic of a subsequent study unit, namely study unit 9.

SELF EVALUATION

(1) Discuss enterprise size in detail. (10)
(2) What are the implication of economies of scale and economies of scope for enterprise size? (10)
(3) Give a detailed definition of a transport market, with particular reference to the homogeneity of a product or service. (8)
(4) Describe the field of study of industrial organisation. (3)
(5) Explain in detail what is meant by short term equilibrium in a market that is characterised by perfect competition and illustrate your answer by referring to the relevant graph. (10)
(6) Discuss short term monopolistic equilibrium in full by referring to the relevant graph. (15)
(7) What is a natural monopoly? (3)
(8) Distinguish between oligopolistic and monopolistic competition. (6)
(9) Can competition be workable to any significant extent if it does not comply with all the requirements for perfect competition? (15)
(10) Discuss the concept of contestable markets and its possible application to transport. (15)
(11) Discuss ruinous competition in full. (15)
(12) Briefly discuss the structure of competition in the markets for the various modes of transport. (15)

STUDY UNIT 8

External costs of transport

CONTENTS

OBJECTIVES

After studying this study unit you should be able to:

- describe the nature of external costs and explain the difference between congestion and pollution
- discuss the different types of pollution associated with transport and the way they are measured
- explain how the optimal pollution charge is determined
- describe the engineering and economic significance of traffic congestion
- explain how to determine the optimal road price and discuss the implications

The costs associated with transport extend further than the private costs of the suppliers of infrastructure and transport services on the one hand, and of the users of transport on the other. The aim of this study unit is to explain the nature of the external costs and their economic implication.

8.1 INTRODUCTION

The full cost of economic activities for the whole of society, in other words the social cost, should be taken into account by the market when the price mechanism is used to effect the optimal allocation of factors of production to the various possible applications. In the previous study units we dealt mainly with the direct costs of suppliers, operators and users of transport. We shall now examine the remaining component of social cost, namely external cost. It is important that you should realise that social cost consists of the direct private costs of the transport suppliers, operators and users. These are the costs we dealt with in study unit 6. But this is not the only element of social cost. Social cost consists of the private costs we have already mentioned plus the external costs. But what are external costs?

You, Mr A, are a resident of a town. Unfortunately your house is on the border of an industrial area. There is a large factory near your house that manufactures products that are daily transported to a market in another province. Large trucks call at the factory both by day and by night and cause a great deal of air pollution and a deafening noise. "If the transport enterprise would pay me some compensation for the noise and pollution I would be able to sell my house, or possibly put up with the situation. But things can't go on like this," complained Mr A to his friend, a member of the town council. "Never mind," said the friend, "I'll think of a way of getting them to pay for the external costs they have caused."

(1) What would you consider the external costs to be?
(2) How could the town council compel the transport enterprise to pay for these costs?

I am certain you identified the air and noise pollution as the external costs. The town council has the option of introducing a noise and air pollution charge which could then have been divided between Mr A and the rest of the community to compensate them for the damages they had suffered. This might possibly lead to a reduction in the volume of transport. At any rate the situation would be more equitable because everyone in the community would be held responsible for the negative impact of their activities.

8.2 THE NATURE OF EXTERNAL COSTS

There are many reasons why the market mechanism cannot always succeed in effecting an optimal allocation of factors of production. In study unit 7 you were introduced to structural market defects that could result in the failure of market prices to reflect the economic costs of products and services. A further reason for the failure of the market is the presence of externalities. Externalities arise when the market attaches no value or price to the beneficial or adverse effects of the economic activities of one entity on another entity or group of entities which are not directly involved. One of the requirements for the existence of a market is that it should be possible to exclude from consumption those who are not prepared to pay for a product or service. The exclusion function does not apply in the presence of externalities, and the result is market failure (malfunctioning). Because the entity responsible for the externality is not remunerated for the external benefits it creates, or is not made to pay for the external costs it is causing for others, there is no reason why it should take the external consequences into account when deciding on the scope of its production or consumption. It will continue to produce or consume at the point where its private marginal utility is equivalent to its private marginal costs. The optimal allocation of scarce factors of production requires, however, that social marginal utility should be equal to social marginal costs. If the external impact of economic activities is "internalised", for example through government intervention in the form of taxes or charges to pay for the external costs caused, and by means of subsidies for the creation of external benefits, and codes of conduct and direct embargoes, it might be possible to move closer to an optimal situation.

In welfare economics we distinguish between different types of externalities. As far as transport is concerned, a useful distinction is that made by Rothenburg (1970) between pollution and congestion. Pollution, essentially, means that some users of a medium abuse it, while other users are the passive victims of this abuse. For example, jet aircraft cause a great deal of noise nuisance, and home owners in the vicinity of airports simply have to put up with this. Road traffic is the classic example of congestion. All the users of the medium (a public road) use it in a similar way, each user impairs the service quality for himself and other users, and the reduction in the quality of the service is approximately the same for everyone. All the members of the user group are therefore equally prejudiced by this self-imposed interaction.

Another important distinction is that between monetary and technological externalities. The difference lies in the fact that technological externalities relate to production or consumption and therefore occur in the cost function (or utility function), which is not the case with monetary externalities. The latter arise when, for example, one enterprise's costs are influenced by price changes as a result of the buying and selling of factors of production by other enterprises. The following example can be used to explain this distinction. The visual encroachment of a new freeway on a view that was formerly unimpaired and pleasant, directly influences the utility function of a resident of the area, and is therefore a technological externality. However, if the same freeway causes the local garage to lose customers to a garage situated next to the freeway, the loss of income by the owner of the former garage is a monetary externality that occurs indirectly, through the price system.

There is a subtle distinction between the two kinds of externalities, mainly because both kinds usually occur simultaneously, but the distinction is nevertheless an important one. Technological externalities do involve factor costs, which have to be taken into account when making decisions involving the optimal use of factors of production. Monetary externalities do not result in factor costs, but they have important implications for the distribution of income – take the example of the profits and losses made by the two garage owners in the previous example. The existence of a monetary externality does not reduce the total benefit of an economic activity, but it does reveal who benefits and who is prejudiced by the activity. The distinction between monetary and technological externalities is therefore extremely important for the evaluation of transport investment by the public sector, where both the scale and the point of impact of the costs are important (Bonsor 1984:20; Button 1982:107–109).

8.3 TRANSPORT AND THE ENVIRONMENT

Transport pollutes the environment in various ways. Mechanised transport causes noise, vibration, poisonous gases, dirt and safety hazards. Furthermore, it often results in the cutting up of residential areas and communities, disruption, loss of privacy and the resettlement of people and industries. Various environment-conscious pressure groups are in favour of the significant reduction and even complete elimination of the environmental consequences of transport, but what they do not take into account is the costs attached to eliminating such disturbances. While some people are forced to suffer the detrimental encroachment of transport in their environment, others notably benefit from the ability to travel and transport goods more freely and cheaply. In virtually all cases the protection of the environment reduces the net benefit to the users of transport. Transport economists therefore tend to think in terms of an optimal reduction in the extent of pollution rather than its total elimination.

8.4 MEASURING POLLUTION

Relatively little is known about the incidence and impact of pollution. In particular, the economic evaluation of pollution causes considerable problems. We shall therefore simply take a brief look at the most important pollution problems.

8.4.1 Noise

Most people, especially in urban areas, find traffic noise a nuisance. One of the most important difficulties about measuring noise pollution in an economically efficient manner is that noise can differ greatly in respect of type, loudness, frequency, duration and timing. Noise affects people in different ways. Some people are sensitive to the slightest noise whereas others appear indifferent even to a deafening row. Background noise is a further problem. As industrialisation and urbanisation advanced, the general level of noise increased as well and with it society's acceptance of a relatively high level of background noise.

The mere development of scales for measuring the level of noise is not sufficient for economists. They have to be able to attach a value to noise in order to calculate the opportunity costs attached to the various noise reduction policies. Various methods of doing this have been devised. One of them is to measure the values of properties at various distances from an airport. The problem with this approach is that property values may differ for a variety of reasons. Nevertheless the approach does provide a practical method of valuation. Another method is to try to establish how much the victims would be prepared to pay for specific reductions in the noise level. This is difficult to establish, however, because individuals have an imperfect perception of reduction in noise levels.

8.4.2 Atmospheric pollution

All forms of mechanised transport cause atmospheric pollution. Even electric rail transport requires energy, which is usually supplied by coal-burning power stations. Atmospheric pollution is an umbrella term used for a whole range of pollutants which are released into the atmosphere in different quantities by different modes of transport. The transport industry emits thousands of tons of solids into the atmosphere annually, with motor vehicles being the biggest culprits. The internal combustion engine is the biggest single source of hydrocarbons and oxides of nitrogen, which are jointly responsible for the formation of smog. In addition to the multidimensional nature of atmospheric pollution, it is difficult to determine the exact toxicity of each substance. Just how hazardous lead oxide is, has been a subject for debate for many years, for example. Human observation of atmospheric pollution is also imperfect, and people associate visible and odorous fumes with toxicity, although this may not necessarily be the case. For example, diesel fumes are less toxic than petrol fumes. Attempts to measure atmospheric pollution usually involve objective estimates of the chemical components of direct exhaust gases.

Although this information offers useful insights, it has little value for the economist. As we mentioned previously, the indirect pollution that is related to the refining of petrol and diesel and the generation of electricity has been left out of account. Besides, data on these sources of pollution are usually too general to form a basis for policy formulation. The level of pollution for any particular mode of transport differs from one vehicle to another and is related to operating conditions. Both these characteristics are very important for policy formulation.

The techniques used to measure the impact of atmospheric pollution are still very primitive. The approach generally used is to measure the number of days lost to production as a result of illness caused by pollution and multiply this by the average value of a day's output. This approach is not very useful because it is too narrow and because of the difficulty in establishing a causal relationship between pollution and illness. People may suffer discomfort as a result of pollution without necessarily becoming ill. Furthermore, this approach takes no account of people's anxieties regarding the possible adverse consequences of pollution.

8.4.3 Traffic accidents

Movement is inherently dangerous. Not only the users of transport but also bystanders such as pedestrians are in danger of being injured or killed. The adverse consequences of traffic accidents are not confined to physical injury to people and property; there is also the fear of possible death and injury that affects the quality of life for people living near airports and busy roads. Because of the difficulties in measuring the fear component, the emphasis falls almost exclusively on the more tangible elements of accident cost.

Although accident statistics give an indication of the relative hazards of different modes of transport, they may not correspond to people's perceptions of accident risk. For example, air transport is safer than road transport all over the world. Nevertheless, people tend to be more nervous about air travel than about road travel. The reason is probably that although road accidents are more common, air crashes are more spectacular.

One method of evaluating the cost of an accident is to take medical costs and the cost of repairs to vehicles into account, as well as loss of amenities of life by the accident victims. The grief and anxiety experienced by friends and relations is sometimes taken into account as well. Death poses special problems, however. There are currently two methods of determining the value of a life that has been saved. The ex post (gross output) method takes into account the loss of production as a result of a death. The difficulty here, however, is that a positive value should really be ascribed to the death of a retired or handicapped person if society would have had to care for them for the rest of their natural lives. The ex ante (net output) approach takes the social benefit of saving a life into account, including the welfare of the person whose life has been saved. This approach therefore concentrates on the additional consumption of products and services that has been made possible by avoiding the death. This approach also falls short, however, because it attaches a financial value only to accidents, and not to the fear of accidents. Recent developments attempt to take anxiety into account as well as death.

8.4.4 Visual intrusion

Transport infrastructure and vehicles are often visually intrusive and aesthetically displeasing. Here again, the problem lies in measuring these consequences of transport. One method is to measure the percentage of the skyline that has been obscured, but this takes only one dimension into account. Transport infrastructure, in particular, should be viewed in the context of its surroundings. A new road located in a formerly unspoilt area would obviously be evaluated differently from a road that partly conceals a waste dump. Design is important as well. It should also be remembered that large vehicles create more disturbance than smaller ones do.

8.4.5 Vibration

Low-flying aircraft, heavy trucks and trains cause vibrations that can have an effect on buildings. Although it is known that the vibrations caused by surface transport vehicles are related to their axle loads, it has not hitherto been possible to relate these vibrations to structural damage. However, there is evidence to the effect that the damage may be less than is sometimes claimed. Improved technology has reduced the damage caused by road transport, and it appears that much of the damage that was alleged to have been caused by heavy trucks was in fact simply triggered by them.

8.4.6 Community severance

Roads, railway lines and other transport arteries often form important physical and psychological impediments to human contact. An urban throughway can cut a local community in two, and in so doing break long-established social ties. It can also make it difficult for people to benefit from job and recreational opportunities on the other side of the barrier. Although it is possible to arrive at an estimate of the delays a traffic artery causes pedestrians, and the time that is lost when destinations have to be changed as a result of the new traffic artery, the suppression of movement is far more difficult to identify. The quantification of community severance is not, therefore, an immediate prospect (Button 1982:111–120).

8.5 THE OPTIMAL POLLUTION CHARGE

We have mentioned that transport economists adopt the approach of trying to keep pollution down to an optimal level rather than eliminate it completely. The optimal point is reached when the marginal social costs of reducing pollution are equal to the marginal social utility. Noise pollution by heavy trucks could be taken as an example. The principles involved are also applicable to other forms of pollution.

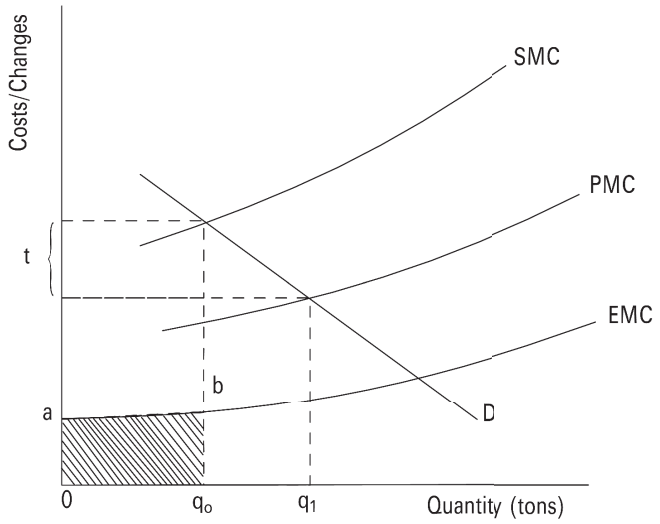
In figure 8.1 PMC represents the private marginal costs attached to transporting goods by lorry in a particular area. EMC represents the external marginal costs, which consist of the monetary value of additional noise at every traffic level. D represents the demand

for road traffic services. The marginal social cost, SMC, is the sum of the private and external marginal costs. If road carriers are unaware of the full cost of their transport activities (or ignore them), they will transport a quantity of goods, q_1, which exceeds the social optimum of q_0, the quantity that would have been transported if all the costs had been taken into account. In terms of the exclusion principle only carriers who are prepared to pay the full external cost of their activities should be allowed to operate in the market. They could be made aware of the external cost of their transport activities by being asked to pay a charge equivalent to the external marginal costs (EMC), that is t at the optimal output level. This would reduce the amount of freight transport to the socially optimal level. In practice this amounts to the transfer of traffic to other modes or the relocation of transport users. This would cause a long-term reduction in total transport inputs.

The charge system poses a few practical problems. Determining the optimal levy or price requires reliable information with regard to the EMC curve. As we mentioned in the previous section, knowledge of the marginal costs of the various kinds of pollution is inadequate and in fact rather primitive at this stage. Furthermore, the income generated by the levy is not directly passed on to the people directly affected by the residual noise pollution. This solution therefore fails to bring about a true improvement in economic welfare. The authority that received the income (equivalent to t $0q_o$) could possibly use it to pay compensation.

FIGURE 8.1

The optimal pollution charge



Source: Adapted from Button (1982:166)

At a theoretical level one could ask whether the exclusion principal is being correctly applied in this case. We are assuming that the hauliers should pay for the right to pollute the environment. The argument could be reversed, however, and we could contend that the "non-hauliers" should pay for the right to an unpolluted environment, that is the hauliers should receive a subsidy equivalent to t to restrict their activities. This question has moral and legal implications involving property rights. The cost of administering taxes and subsidies should also be taken into account.

A further problem with charging polluters is that our imperfect knowledge of external marginal costs makes it impossible to judge whether the charge imposed is too high or too low, even if an arbitrary amount is charged at first. The usual trial-and-error method of determining prices in the market can therefore not be used. But since any optimal containment of noise, irrespective of the method used, requires knowledge of the external marginal costs, a charge system could just as well be introduced because it would probably cause the least distortion in the market. This means that the cost of keeping the environment in an acceptable condition should be reflected in the cost of the products and services that result in pollution, during either the production or the consumption phase (Button 1982:166–168).

8.6 CONGESTION

Although transport capacity can be adapted to changes in demand over the long term, it is fixed in the short term and cannot be adapted to daily, seasonal or other changes in the demand volume. The capacity of an airport, for example, cannot be expanded or reduced to meet seasonal demand. When the users of a particular facility start disturbing one another owing to the capacity of the facility, we speak of congestion. Some degree of congestion is unavoidable if we wish to avoid serious underutilisation of the facility, but the question is, how much congestion is acceptable. Because transport users will accept some level of congestion, but will resent excessive congestion on account of the time and inconvenience cost involved, the concept of an optimal level of congestion has evolved. This concept can best be explained once we understand more about the nature and underlying economic causes of congestion.

The root cause of congestion is that a number of users have to share the service of a facility with a fixed capacity. As the number of users increases, they begin to interact negatively with one another, which affect everyone in a similar way. Additional users reduce the level of service of the facility to virtually the same extent for themselves and for other users. The result is that the utility a user derives from his or her use of a facility (or the value he or she attaches to the facility) is directly influenced by the consumption behaviour of other users.

Generalised costs supply the necessary link between physical traffic flow and cost. Generally speaking, rapid transport implies cost saving. Vehicles are more efficiently used and transport or travelling time is saved. However, increases in traffic density cause decreases in average travelling speed, with a consequent increase in the generalised costs for all the road users concerned. The economic reason for the development of traffic congestion is that additional motorists who join the traffic are either ignorant of the full costs they are causing or unwilling to take them into account in their decision making. The result of this problem is that the optimum economic use is not made of the fixed infrastructure. The transport economist is not primarily interested in the transport engineering aspects of the problem of congestion. What is of greater importance to the transport economist is that congestion leads to the less-than-optimal use of scarce resources. The transport economist's task is to find a professional solution to the problem of making the users of a road on which congestion occurs pay their full costs. Remember that congestion is an external cost and that if that cost is not included in the social cost then we have not achieved the maximisation of social welfare. How should the transport economist make provision for congestion costs? He or she would do so by means of a congestion charge. We shall go on to discuss this.

8.7 CONGESTION CHARGE

The optimal use of factors of production requires that road users pay the full cost of their road usage or else be excluded from the use of the road (naturally this applies to all transport facilities and services). A congestion charge affords a method of getting road users to pay for the full cost of their road usage. A congestion charge should be distinguished from other forms of charging paid by road users, such as license fees and excise tax, which bear little relation to road usage. Congestion levies are specifically related to the level of congestion costs generated by road users.

It is sometimes contended that a congestion charge is an undesirable method of reducing excessive congestion, and that direct control in the form of permits or embargoes is preferable. The argument is that road space should preferably be allocated by means of quotas. The problem about such an approach is that there are no clear-cut criteria to apply to the issue of permits. Road users who should have free access can easily be identified; these would include doctors, emergency services, the handicapped and the police, but these people account for only a small percentage of all road users. Urban commuters, in particular, pose a problem because there are no objective criteria for determining whether or not they should be allowed to take their cars into the city.

The way the optimal congestion charge is determined can be illustrated with the aid of a simple algebraic model and figure 8.3. Suppose that at a flow rate of n vehicles per hour the traffic flow is unimpeded and the average general cost is equivalent to c. When an additional vehicle joins the traffic stream the average travelling speed decreases and operating costs increase, with a consequent increase in the average generalised cost. Traffic flow now increases to n + 1 vehicles and the average generalised cost increases by c to c + c for all the road users, including the additional user. The additional user's private marginal cost (PMC) therefore amounts to c + c, which is the same as the social average cost (SAC) of all the road users. The additional cost for the entire traffic stream, which consists of the private costs of the additional road user and the cost he has caused the other road users (the social marginal cost or SMC) is equal to:

$(n + 1)(c + \triangle c) - nc = c + \triangle c + n\triangle c$

The SMC (which is also the external cost) is therefore $n\triangle c$ larger than the PMC of the additional road user. Because PMC is lower than SMC, there is a natural tendency for road usage to exceed the social optimum.

FIGURE 8.2

The congestion charge

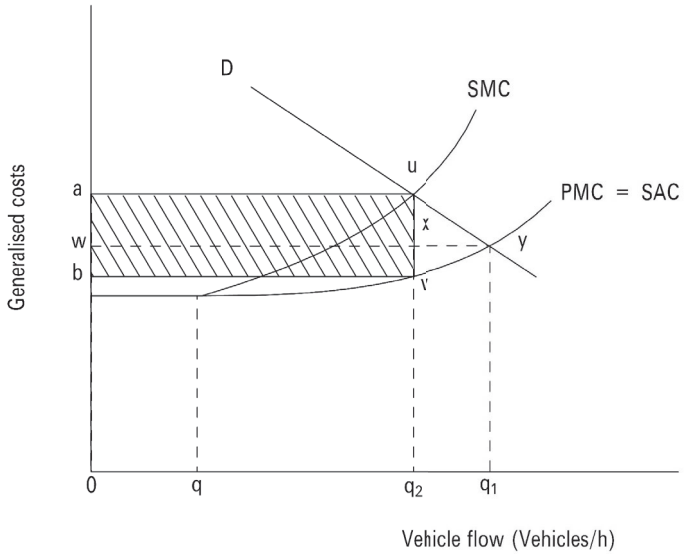


Figure 8.2 shows that at a flow rate of q vehicles per hour there is no congestion. If there is any further increase in the flow rate the vehicles begin to impede one another and SMC begins rising more rapidly than PMC. Because the road users only take their own (private) costs into account, equilibrium is reached at y, where the demand for road space D and PMC are equal. At this point the traffic flow is equal to q_1 and the generalised cost is equal to w. The traffic flow q_1 is higher than the socially optimal traffic flow q_2, which would have occurred if social marginal costs (SMC) and marginal utility (represented by the curve D) at u had been equal. This means that in order to affect socially optimal traffic flow an additional road user should, in addition to paying his private costs, also pay the external costs of his road use. This optimal road price is represented by ab in the figure.

After the introduction of a road charge the traffic flow decreased by $q_1–q_2$, and the road users who left the road lost a user surplus equivalent to the area uxy. At the same time the authorities collected an amount equivalent to the area auvb, that is an amount equivalent to the product of the road price ab and the traffic flow Oq_2 (which is the same as bv). However, this amount does not represent a net social benefit, because a portion of the amount (equivalent to the area auxw) represents a consumer surplus transferred from the remaining road users to the government. The net social benefit therefore amounts to wxvb minus uxy. One shortcoming of the system is that it is the authorities that derive the benefit from the road charge; however this benefit could be returned to the road users by means of a distribution system of one kind or another.

The practical implementation of the payment of road charges presents many problems. Progress in the field of electronic technology has, however, made the introduction of experimental systems possible.

8.8 CONCLUSION

If the optimal utilisation of scarce factors of production is to be achieved, the external cost of transport must be included in the price. The measurement and economic evaluation of external costs, and particularly of pollution costs, and the implementation of practical systems of payment, will still require a great deal of research.

SELF EVALUATION

(1) What is an externality? (3)
(2) Distinguish between pollution and congestion, and between monetary and technological external costs. (6)
(3) Discuss the effect of transport on the environment. (10)
(4) What is the most important limitation as regards the introduction of pollution charges? Briefly discuss this question with reference to various kinds of pollution. (20)
(5) Explain the determination of the optimal pollution charge with the aid of the relevant curves. (15)
(6) Explain the concept of congestion. (5)
(7) Explain how the congestion charge is determined. Illustrate your explanation by both algebraic and graphic means. (30)

STUDY UNIT 9

Pricing in transport

CONTENTS

OBJECTIVES

After having studied this study unit you should be able to:

- explain the economic reasons for and the nature of marginal cost pricing
- discuss the reasons for and the nature of next-best pricing
- define and discuss price discrimination in detail
- explain pricing with reference to joint costs in detail

Various factors – in relation to both supply and demand – complicate the pricing of transport services. The aim of this study unit is to put some of these problems in perspective from the point of view of welfare economics.

9.1 GENERAL

In a market system prices fulfil an important function in the allocation of scarce factors or means of production among the various applications. The price charged for a product or service reflects the objectives which the particular entrepreneur or enterprise is pursuing. For example, a price aimed at maximising profit may differ from a price aimed at maximising sales or promoting general welfare. Some enterprises are not trying to maximise or minimise anything, but are pursuing a pricing policy aimed at attaining certain lower order objectives, such as the security of the enterprise or the maintenance of a minimum market share. Generally speaking, prices may be intended to promote the welfare of the enterprise (usually the case with private enterprises) or the welfare of the consumers (usually the case with public enterprises).

In this study unit we have approached pricing under various market conditions from a welfare economics perspective. This means that prices are determined with the aim of promoting social welfare, as opposed to the approach in business economics, where prices are determined with the aim of maximising entrepreneurial profit. Under conditions of perfect competition these two objectives should coincide and social welfare is maximised when private enterprises maximise their profits.

Before we continue with our discussion of pricing, there is one point that it is important to get clear. In economic theory we distinguish between positive or descriptive economics, the object of which is simply to describe and explain the operation of the economic system, and normative or welfare economics, which is prescriptive with regard to economic policy. Because the government has always played an important part in transport, through means such as the regulation of transport prices, and because transport is often used as a policy instrument by the government, the dictates of welfare economics have always had an important influence on transport pricing.

9.2 ECONOMIC EFFICIENCY

Economic efficiency is a key concept in welfare economics. The allocation of scarce factors of production is the most efficient in the economic sense (known as the Paretian optimum) when it is such that no consumer could benefit from a reallocation of factors of production without another consumer being adversely affected. In a situation of this kind there is maximum satisfaction of needs. In a free market system the choice of what should be produced with the available factors of production should be left to the market participants. In order to enable consumers to make the choices that will lead to the maximum satisfaction of needs, the prices of goods and services should reflect their economic or opportunity costs. This enables consumers to decide whether the degree of utility or satisfaction of needs they derive from using a product or service makes up for going without other products or services.

It may well be asked which type of cost should be used as a basis for pricing. The demand for goods and services is usually sensitive to price, in other words, it is price-elastic. If a consumer has to decide whether to buy a little less or a little more of a particular product or service, its price in relation to the prices of all the other competitive products and services should reflect the cost of producing a little more or a little less of that product, in other words its economic marginal cost. If a consumer has to pay more than the marginal cost, for example the price charged by a monopoly, he would buy less than the socially optimal quantity because the price exceeds the marginal costs. If the price is lower than the marginal costs, the consumer would buy more than the optimal amount.

The most important implication for a policy aimed at optimising economic efficiency is therefore that prices should be equivalent to marginal costs. The price of a product or service reflects the marginal income of the producer or seller, and the above policy directive therefore means that marginal costs and marginal income should be equal to one another under conditions of perfect competition, where marginal income, average income and price are equal.

There are two further requirements for economic efficiency.

(1) Prices should reflect the total marginal cost of production and consumption. This would include both private and external costs.
(2) The prices of all products and services in the market should be equivalent to their marginal cost, otherwise the "problem of the second best" arises.

In perfectly competitive markets, the market mechanism ensures that prices are equivalent to marginal cost, which in turn brings about an economically efficient (Paretian optimum) allocation of scarce factors of production.

Although economic efficiency clearly requires that prices and marginal costs should be equal, it is not clear as yet whether long-term marginal costs (LMCs) or short-term marginal costs (SMCs) are intended. We shall now take a look at the time aspect of marginal cost pricing.

9.3 THE TIME PERSPECTIVE OF MARGINAL COST PRICING

The economic premise that sunk costs are irrelevant to decision making on prices and production was discussed in study unit 6.

If capacity is accepted as a given, only the variable or avoidable costs of production or operations can be affected by these decisions.

The proportion of variable or avoidable costs to total costs increases, however, as the period under consideration increases. The longer the existing plant and equipment have been operating, the higher the variable costs associated with shutdowns, repairs, maintenance and wastage of raw materials and labour. When the above variable costs have increased sufficiently, the replacement of the plant and equipment would be considered. At this stage a decision should be taken with regard to fixed or capital costs, which could mean that a cost that had been fixed in the short term becomes variable. The question now arises whether marginal cost pricing should include only the short-term costs of nondurable means of production, or the medium-term costs, which would include the cost of repairs or maintenance, or possibly long-term costs, which make provision for the renewal, replacement and expansion of durable means of production.

The relevant economic principles are relatively simple, but their application is not by any means simple:

(1) First, marginal costs represent only those costs that are causally related to the production of the additional output. Buyers are required to pay only for the cost of producing the additional unit in question.
(2) Secondly, only short-term costs are involved.

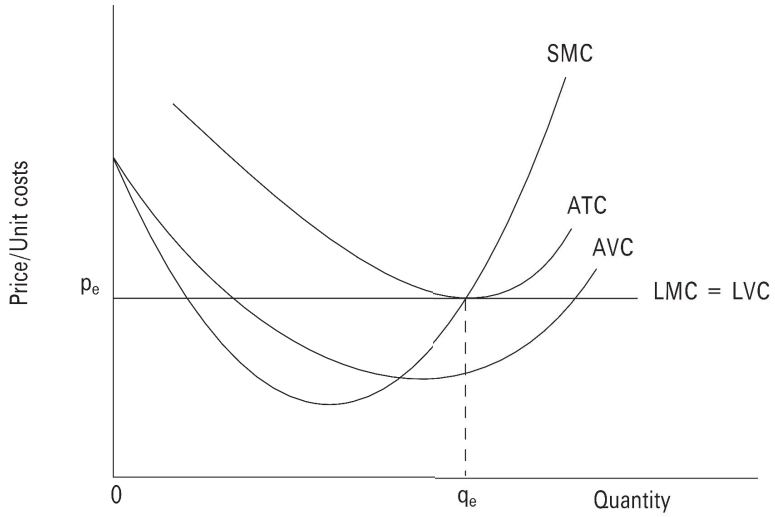
At first sight, these two principles appear to contradict one another. The first implies that all costs should be taken into account, irrespective of when they were incurred. If additional production leads to an increase in repair and maintenance costs, for example, it should include marginal costs, even if this increase will only be incurred at a later date.

The second principle appears to exclude these costs, which only become an issue in the long term. According to Kahn (1971:71), the sacrifice of future value or any future costs that are causally related to current production should form part of the variable cost of additional production, also referred to as usage costs. In terms of this approach, short-term marginal costs are therefore equivalent to the change in the total variable cost of producing the additional unit. In that the technical wear and tear on a plant or equipment is linked to use, replacement costs constitute part of variable costs. If the price of a product or service does not cover this cost, it is not a reflection of the total marginal cost of the additional unit.

The second principle therefore does not conflict with the first: Only those costs that are causally related to the production of the additional unit, irrespective of whether they will be incurred immediately or only at a later stage, are taken into account. It therefore follows that inasmuch as maintenance and replacement costs and the cost of capital are not a function of use, they do not form part of the short-term marginal costs and therefore not of the price. The part of the operating cost of the plant or equipment that is affected by use should be valued at replacement value, since the increase in production is responsible for a future cost or a decline in the future value of the plant and equipment (and not its historical cost). The increased repair and maintenance costs should only be incorporated in the marginal costs if they are likely to be incurred, in other words if the anticipated demand will be sufficient to justify the additional expenditure.

FIGURE 9.1

The unit costs of an enterprise with constant long-term unit costs



Suppose the only possible present or future use of the plant or equipment in question is the production of additional output, and that the demand is such that in the foreseeable future the price will not be able to cover the additional replacement, maintenance and repair costs that will result from a decision to produce (rather than not to produce). As long as buyers are prepared to pay a price that covers the immediate variable operating costs, including a return on the salvage value of the plant and equipment, it would be more advantageous for both the owners and society to accept that price and continue to operate rather than to shut down. Operations should be continued until the higher operating costs are no longer covered by the price or the necessary repairs, maintenance and replacement costs can no longer be postponed.

The second principle apparently raises another difficulty. It appears that prices based on short-term marginal costs will never cover fixed costs, such as replacement value and return on capital. This is not necessarily the case, as is evident from figure

9.1. Although the short-term variable cost per unit of output (AVC) can never be as high as average total costs (ATC), short-term marginal costs (SMC) may be less than, equal to (at p_e) or higher than ATC, depending on the market demand and the capacity of the industry. At outputs larger than q_e, SMC rises to above ATC and prices based on short-term marginal costs (SMC) cover more than ATC. Pricing on the basis of short-term marginal costs need therefore not be unprofitable in the long run or inconsistent with long-term equilibrium. The price need not be explicitly set to cover long-term or fixed costs, but it will cover these costs at times when demand is sufficiently high.

The ideal, briefly, would be to set all prices at short-term marginal cost, with appropriate adjustments for the second-best situation (see section 9.8 and Kahn 1971:70–75).

However, there are several possible objections to using short-term marginal costs as the basis for prices. What happens, for example, if prices based only on variable operating costs cause so large a demand that existing capacity is insufficient to satisfy it? (Congestion levies were discussed in study unit 8 and are one of the possible solutions.) Apart from possible congestion, there are other reasons why short-term marginal costs cannot be a suitable basis for price.

(1) This can cause an unacceptable price fluctuation where the relationship between demand and capacity varies over time.
(2) It can be difficult, frustrating or expensive to calculate the changing congestion or opportunity costs mentioned in (1) and base the price on them.
(3) Prices calculated on this basis may not cover the average total costs of the facility (eg a bridge) over its useful life, and a public subsidy may be necessary.

If a decision is taken to base prices on long-term marginal costs (LMC), provision should also be made for the fact that additional demand may require additional capacity at certain times. This means that capital or capacity costs should be included in the price. It therefore appears that in particular circumstances, such as where capacity costs amount to a large portion of total costs and the facility in question is subject to major fluctuations in demand, the use of LMC as a basis for prices could be a practical solution.

9.4 SPECIFYING THE UNIT OF OUTPUT

The magnitude of the incremental cost naturally depends on the increase (increment) in output. If an aircraft on a scheduled flight is ready to depart and there are vacant seats available on the aircraft, the marginal or incremental cost of carrying an additional passenger would be virtually nil. But flight schedules can be changed in the medium term, and when this happens the cost of the entire flight is relevant. If the incremental output calls for a regular service between two cities, still more additional costs have to be taken into account. The larger the incremental "unit of output", the greater the number of costs that become variable, so that even the purchase of aircraft and overheads such as office rental can be included under incremental costs.

The choice of an incremental output unit, like the choice between the long and the short term, involves weighing up the economically ideal solution against what is feasible in practice. The ideal is the smallest possible unit at marginal cost. However, the appropriate size depends on the nature and impact of the decision under consideration. From the transport operator's point of view a decision to add another destination to a particular route involves a different marginal cost (which has to be compared to marginal income) from the cost involved in a decision to include an additional flight or route in his schedule. When the decision concerns the withdrawal of a service the cost of the entire service has to be taken into account. The relevant economic question is therefore what additional costs would arise if a service were to be introduced or expanded, or what costs could be saved or avoided by curtailing or withdrawing a service (Kahn 1971:75–77).

9.5 SECOND-BEST PRICING

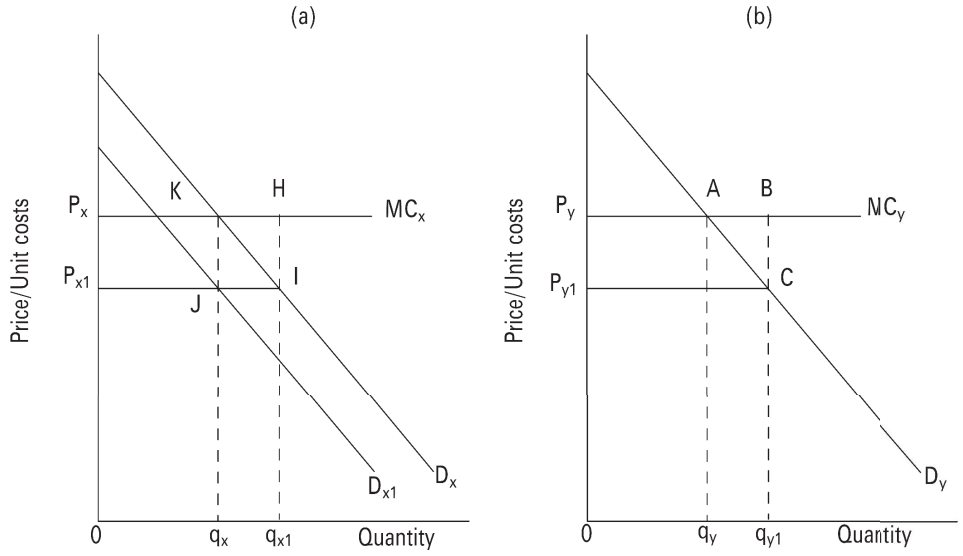
In section 9.2 we mentioned that one of the requirements for economic efficiency is that the prices of all products and services in the market should be equal to their marginal cost. Any factor that causes a price in the market to deviate from marginal costs hampers the optimal allocation of factors of production. The question that arises is: If any deviation from marginal cost pricing occurs in the market, would pricing the remaining products or services on the basis of marginal costs lead to a second-best solution? According to the theory of the second best, which was formally summarised by Lipsey and Lancaster for the first time in 1956, the answer is "no". This means that, as we shall explain shortly, a deviation in marginal cost pricing means that all products or services should no longer be priced at marginal cost.

Suppose there are three components in a country's economy: One (Z) in which prices are equal to marginal costs, a second (X) in which all prices are above or below marginal cost, and a third (Y) in which the government fixes prices. (The prices in Z are already optimal and will therefore not be discussed any further. They are accordingly not shown in figure 9.2.) Suppose, further, that X is a public road in respect of which the price is less than the marginal cost because external costs have not been included, and that Y is a railway enterprise where prices are equivalent to marginal costs. In figure 9.2(a) p_x is the optimal road price and

p_{x1} the price that is in fact being charged. Consequently the quantity of X that is consumed is q_{x1}, which is more than the optimal quantity q_x. The demand curve for road transport D_x is based on the assumption that rail tariffs are equivalent to marginal cost and that the demand for rail transport is represented by D_y in figure 9.2(b). In the case of rail transport these assumptions yield the optimal combination of prices and quantity, namely p_y and q_y, as represented in figure 9.2(b). If rail tariffs are reduced to below marginal cost (remember that the road price p_{x1} is below marginal cost), that is to p_{y1}, the demand for road transport declines from D_x to D_{x1} because the two are substitutable. Although the road price does not change, the use of X declines from q_{x1} to q_x, and the use of Y increases to q_{y1}.

FIGURE 9.2

Second-best pricing



We are now able to study the effect on economic welfare if prices were fixed below marginal costs in Y. The increase in total factor costs from q_y to q_{y1} owing to the increase in demand is equivalent to the area q_yABq_{y1}, whereas the total increase in user benefits is represented by the area q_yACq_{y1}. The net economic welfare is therefore reduced by the area of the triangle ABC. In the road industry (X) the decline in production leads to a saving in the quantity of factors of production that is equivalent to the area q_xKHq_{x1}, and the reduction in user benefits is equal to q_xJIq_{x1}, which gives us a net increase in welfare equal to KHIJ. Provided KHIJ is greater than ABC, total economic welfare will be improved by below marginal cost pricing in the controlled area of the economy (Y). This is precisely what is meant by a second best optimum. It does not mean that welfare is optimised in all sectors of the economy as in the case of marginal pricing; rather it is an attempt to improve a suboptimal situation.

The above discussion brings us to three basic guidelines on second best pricing:

(1) If prices in the distorted market are set below marginal cost, the prices in the controlled market should be set below marginal cost if the products or services are substitutes, and above marginal cost if they are complements.
(2) If prices in the distorted market have been set above marginal cost, the prices in the controlled market should be fixed above marginal cost if the products or services are substitutes, and below marginal cost if they are complements.
(3) If the products or services are neither substitutes nor complements, the prices in the controlled market should be set at a level equal to marginal cost (Bonsor 1984:26–28).

The existence of monopoly power, external cost and other distorting influences in the economy necessitates certain adjustments to marginal cost pricing in the transport sector. The above guidelines are based on a simple example of two industries each of which produces a single product. Reality is far more complex, however, and even the data obtained from distorted markets do not offer a reliable basis for calculating optimal deviations from optimal cost pricing. These guidelines, in conjunction with the dictates of economic efficiency, are no more than useful pointers in the formulation of economic policy. Each suboptimal situation has to be assessed on merit.

9.6 PRICING IN THE TRANSPORT SECTOR

It is important to begin by indicating the three types or methods of pricing that could be adopted by transport enterprises. These three types or methods hinge around the basis on which tariffs for transport services are determined. They are

- cost-oriented pricing
- demand-oriented pricing
- competition-oriented pricing

9.6.1 Cost-oriented pricing

In this case the cost is determined first and a tariff or price is subsequently determined on the basis of the cost. Depending on which cost is used as the basis for pricing, we can speak of total cost prices, marginal cost pricing et cetera. A fixed profit percentage is usually added to the cost and this forms the price requested. In the long term the price must cover costs. This method has a number of benefits:

- It is a method that can easily be applied and it need not change as the demand changes.
- If all the transport entrepreneurs use this method there is no unnecessary price competition.
- From a welfare perspective, if prices are based on marginal costs, in particular, a price that is fairer to both the suppliers and the users of transport can be arrived at.

9.6.2 Demand-oriented pricing

Whereas cost was the determinant for price in the previous case, demand is the determinant here. Demand will be intensively examined. It is demand elasticity that will determine what price will be demanded for a particular group of consumers. The transport enterprise may not see any differences between related products or services on offer, but the consumer views them in a different light. The consumer is therefore prepared to pay different tariffs, depending on his elasticity of demand. If these differences are based on cost differences (a particular service is a little more expensive because it requires a better quality) we speak of price differentiation. But if the costs of the two services are the same but the different price elasticities of demand allow the entrepreneur to charge different prices, we speak of price discrimination. See the following section (section 9.7) on price discrimination.

9.6.3 Competition-oriented pricing

This method is not based on cost or demand, but is a response to the requirement to remain competitive. A transport enterprise may decide, for example, that in order to penetrate a new market segment it will charge a price just below the generally applicable price in the market. The assumption is usually that the market participants have already calculated the cost and that the price they are charging covers costs and ensures a profit. The new entrant, who is trying to lure their clients away, is prepared to take a smaller profit.

Later on the new entrant would raise his prices a bit and allow him to be guided by the generally prevailing market price.

9.7 PRICE DISCRIMINATION

9.7.1 General

Price discrimination is practised when the same commodity, or variations of the same commodity, (product or service) is sold at different net prices to different consumers. By net price is meant the selling price less any cost involved in serving a particular consumer, such as transport costs or the cost attached to the changes in the product specification. Price differences due to the inclusion of cost items such as the latter do not really amount to price discrimination. Price discrimination is obviously a deviation from the welfare dictate of marginal cost pricing.

With multiproduct enterprises such as transport enterprises we should remember the problem of joint costs and joint production. This means that it is more difficult to keep strictly to the definition of price discrimination. Strictly speaking, any price that differs from another price although the average costs are the same is a case of price discrimination. But what happens if there are joint costs that will only be allocated after the price has been determined? You are referred back to study unit 6 of the study guide, which deals with this problem. There is no unique cost basis that determines whether or not prices should be regarded as discriminatory.

This creates a difficult situation for the transport entrepreneur. He or she wishes to fix the prices for the various services in a manner that will maximise profits and does not want joint production to weaken the profit position.

There are a number of requirements to be met before price discrimination can be applied. These requirements, which are related to the degree of monopoly power exercised by the seller and the price elasticity of the demand of his various buyers (submarkets) are as follows:

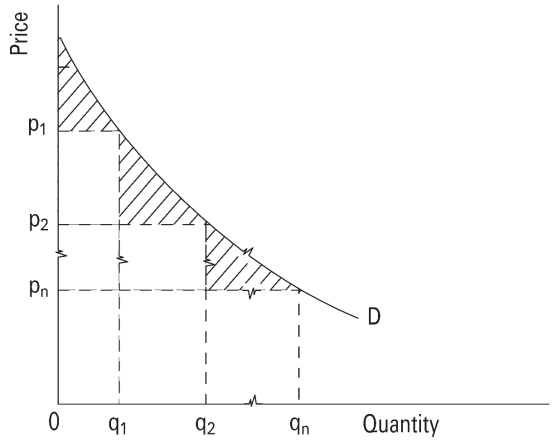
(1) Firstly, it should be possible to separate the various submarkets. Price discrimination is impossible if demand can be transferred from one market to another. It is important in this regard to distinguish between the transfer of a unit of the commodity (arbitrage) and the transfer of a unit of demand.

Let us first consider the transferability of commodities. Certain services have a direct bearing on the buyer or client and are completely untransferable. An example here is the different fees charged by doctors, lawyers and dentists in accordance with the income level of their clients. The same principle applies when the service has a direct bearing on particular commodities, such as transport services. If a transport enterprise charges lower tariffs for coal handlers than for copper handlers, nobody can change their copper into coal for transport purposes in order to take advantage of the lower tariff.

Price discrimination becomes more lucrative as the possibility of transferring the unit of demand becomes smaller. The highest degree of untransferability occurs when the separation of buyers is based on the personal attributes of buyers. Wealthy people do not suddenly become poor, for instance, simply in order to qualify for legal aid. Some personal attributes, such as differences in taste or time preferences, do not preclude a transfer to cheaper markets. In such cases the seller would create untransferability by using special trademarks, special packaging for different varieties of a commodity and advertising campaigns. In particular, it may be necessary to enlarge the quality differences between higher-priced and lower-priced goods or services in order to prevent buyers who could afford the higher prices from switching to cheaper versions of the commodity. In this respect the transferability of demand does accentuate the need for discrimination.

FIGURE 9.3

Second-degree price discrimination



(2) Secondly, it should be possible to differentiate between markets, in other words it should be possible to identify groups of buyers with different price elasticities of demand. This is especially important where a degree of transferability of demand is present. It can be a very expensive process to gather the information necessary to distinguish between different clients. This explains the popularity of self-selecting devices. Such systems enable buyers to decide for themselves which markets they would like to buy in, and what quantities they would like to buy. The choices available between cheaper season tickets or more expensive single tickets for an opera season or between different classes of airline tickets enable them to choose for themselves in which submarket they wish to buy and what quantities they want to buy.

(3) The final requirement is a logical consequence of the previous one. The market of a seller with monopoly power who wishes to discriminate must necessarily contain buyers with varying price elasticities of demand.

Pigou (1920) was the pioneer in distinguishing between three kinds of price discrimination. Each of the three kinds is related to the degree of monopoly power enjoyed by the seller:

(1) First-degree or perfect discrimination involves setting different prices for each unit of a commodity that is sold in such a way that each price corresponds exactly to the demand price and the full user surplus is transferred to the seller. The demand price, which is also known as a reserve price, is the highest price the buyer would be prepared to pay. One method of bringing about first-degree discrimination would be to negotiate independently with each individual consumer, but this is time-consuming and expensive. This kind of discrimination is also seriously hampered by the publication of prices or tariffs.

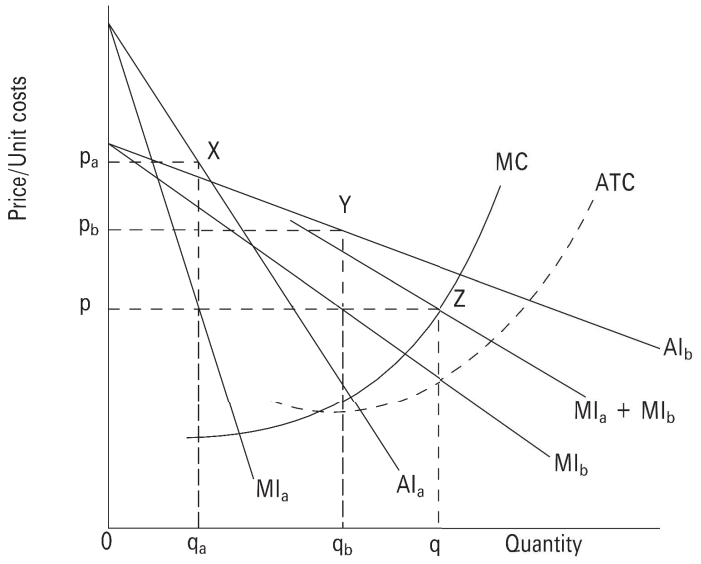
(2) Second-degree discrimination is encountered when an enterprise is able to charge n different prices in such a way that all units with a reserve price greater than pp are sold at p_1, all with a reserve price greater than p_2 at p_2, et cetera until p_n is reached. Consumers are therefore placed in groups, the members of which all pay the same price. As shown in figure 7.3, some consumers still retain some of their user surplus (the shaded areas) because their reserve price is higher than the price for the particular group.

    Note that all the buyers are able to buy whatever they wish at a price that is equivalent to or lower than their reserve prices. First-class, second-class or third-class train tickets are a case in point, if we accept that all buyers are able to afford at least a third-class ticket. Second-degree discrimination often makes use of self-selection systems.

(3) Third-degree discrimination is similar to second-degree discrimination and is found when a monopoly is able to identify a number of submarkets on the basis of a useful attribute of one kind or another and obtain a separate monopoly price for each separate subgroup. It differs from the previous two forms of discrimination in this respect that the monopolist can refuse to sell in one market at reserve prices that would have been adequate in the other markets. An example of spatial pricing serves as an illustration.

FIGURE 9.4

Price discrimination by a private monopolist



Suppose a monopolist sells a product in three separate geographical markets all of which are situated next to a particular road. The cost of a return ticket between any two markets is R10. The enterprise charges the following delivered prices (which include transport costs) in each market: First market – R90, second market – R95, third market – R100. A buyer in the third market whose reserve price is R99 is unable to buy in that market, or in the other two. The total cost of buying in the second and first markets would be R95 + R10 = R105 and R90 + R20 = R110 respectively. The same applies to buyers in the first and second markets (Philips 1983:1–19). Third-degree discrimination is commonly practised in the transport sector and the demand elasticity of various submarkets can be used to distinguish between markets. This form of discrimination is discussed in the next section.

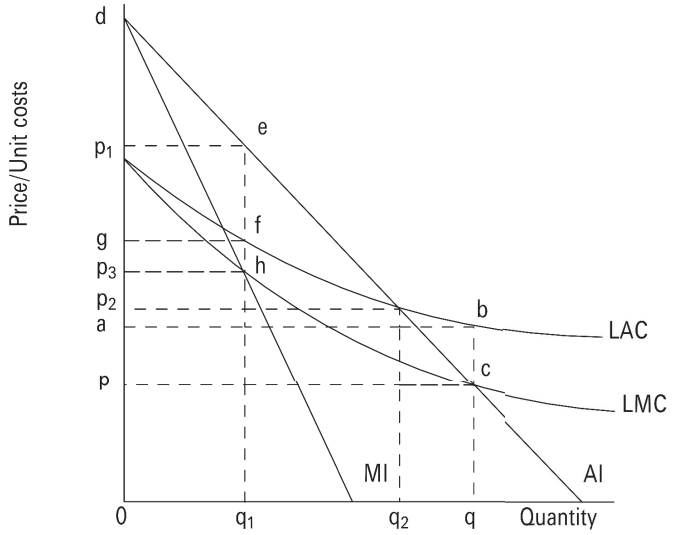
Discriminatory pricing usually reflects opportunities to earn a higher income in different submarkets in order to realise a financial objective of one kind or another, such as cost recovery or profit maximisation. This means that the monopolist must take into account such factors as differences in time and location, and profit differentiation, when determining prices. A transport operator is able to differentiate between submarkets on the basis of different kinds of user, the goods transported, the particular route served, the geographical location of users, the timing of the demand and the by now familiar quality attributes of transport.

Figure 9.4 illustrates a situation in which a private monopolist markets his product or service in two different markets, A and B. The total marginal income curve, $MI_a + MI_b$, is obtained by means of a horizontal summation of the marginal income curves of the two separate submarkets. This curve shows the maximum marginal income that could be earned by selling an additional unit of the service in question. The output q at which the monopolist maximises his profit is determined by the cut-off point of the marginal income curve MC (and because only one product or service is supplied, there is only one marginal income curve) and the total marginal income curve. The enterprise maximises its profit when its output is divided between A and B in such a way that the marginal income of the last unit sold in each of the markets is equal, that is when $MI_a = MI_b$. If MI_a should be greater than MI_b, it would pay the carrier to sell more in A and less in B. Figure 8.4 shows that the carrier sells a quantity q_a at a price p_a in market A and a quantity q_b at a price p_b in market B, with $q_a + q_b = q_c$. It therefore appears that the price (p_b) in the market with a higher price elasticity (B) is lower than the price (p_a) in the market with the lower price elasticity (A).

According to the welfare principle of economic efficiency, price should be equal to marginal cost. A public enterprise that sets its prices in accordance with this standard would still cover its costs in a case such as that illustrated in figure 8.4, because marginal costs lie above average total costs (ATC) when the total average costs rise, and the price p is sufficient to cover ATC.

FIGURE 9.5

Pricing in the presence of decreasing long-term unit costs



But when long-term unit costs are declining, however, the situation differs considerably. As shown in figure 9.5, the demand curve (AI) and the marginal income curve (MI) intersect the long-term average cost curve (LAC) and the long-term marginal cost curve (LMC) where they decline. The social optimal output and price are q and p respectively, in other words where LMC and AI intersect, since at this stage the monopolist is no longer making an excess profit. A profit-maximising monopolist would not produce at this point, because the price p falls below LMC, and he would consequently suffer a loss equal to pabc (or bc per unit). He would fix his price and output at p_1 and q_1 respectively, which would give him a monopoly profit of p_1efg. In the case of a railway enterprise this would also mean that the quantity q to q_1 would not be conveyed.

The problem now is to get the monopolist to produce at a level that is closer to the social optimum while at the same time preventing him from suffering losses. Simply compelling a railway enterprise to break even, in other words to operate where LMC = AI with an output of q_2 and at a price p_2, is not a good solution, because a quantity of traffic (q to q_2) would not be conveyed.

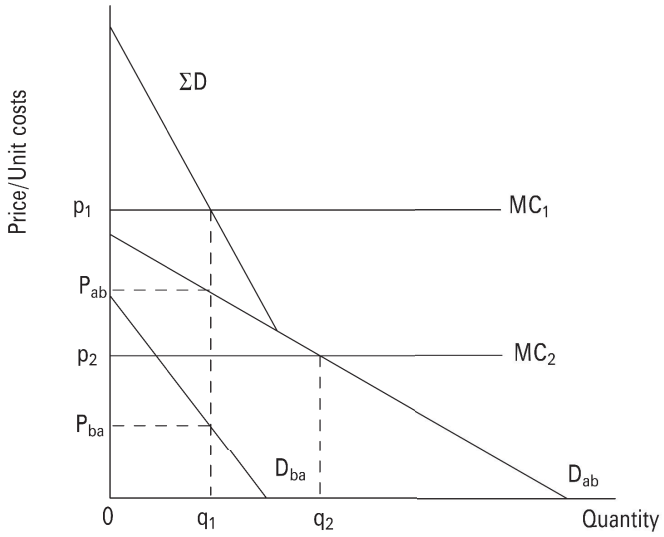
Another solution would be to apply marginal cost pricing, in other words an output of q_1 at a price p_3, and make the loss good out of general taxation, but apart from the fact that this involves an undesirable transfer of income from the general taxpayer to the consumers, the output is still q to q_1 less than the social optimum. The key to the solution lies in price discrimination. If output q is produced at the social optimum, the consumers enjoy a surplus of dcp, which can be used to offset the loss of pabc. Although perfect discrimination is not possible, the public enterprise (in this case a railway enterprise) can discriminate in the same way as the profit-maximising private monopolist in figure 9.4 between submarkets with different price elasticities of demand. This method is one of the variants of value-of-service pricing (which we also previously defined as demand-oriented pricing), and because information regarding the price elasticity of demand of commodities is generally not available, it is generally accepted that price elasticity is inversely proportional to unit value.

Commodities with high values are therefore transported at higher tariffs than commodities with low values. Since a public transport enterprise's aim is merely to break even, it would probably not be necessary to carry price discrimination to the extreme.

As you already know, decreasing long-term unit costs are found, among others, in enterprises with high fixed costs as a result of indivisibilities, such as railway enterprises and other public utilities. According to Kahn (1971:133), price discrimination gives enterprises of that nature the opportunity to (1) cover average total costs (or come closer to doing so), while (2) making fuller use of their existing capacity, and (3) taking fuller advantage of long-term decreasing unit costs by (4) selling as much as possible as long as buyers are willing to bear the incremental (marginal) cost of supplying the service to them.

FIGURE 9.6

Pricing in the presence of joint costs



9.8 PRICING WITH A HIGH PROPORTION OF JOINT COSTS

The problem of choosing an appropriate output unit is only one aspect of the more general problem surrounding common and joint costs. The cost of carrying fifty passengers between two cities by bus or aeroplane can be expressed as single figure which represents the joint cost of carrying all the passengers, and not costs on an individual passenger basis. The marginal cost of providing an additional sales unit (seat) is therefore virtually nil, whereas marginal production or operating costs (offering an air or bus transport facility available between the two cities) can amount to a considerable sum. The existence of common and joint costs means that it is impossible to link an identifiable cost to individual sales units of products or services (revise section 4.3.3).

Joint costs pose very few practical problems in the market. If a road transport service that offers a weekly return trip between A and B with M trucks is offered in a competitive market, equilibrium rates would soon emerge for the two hauls (from A to B and from B to A). Although a specific cost is attached to the picking up and handling of consignments, it makes very little difference to the total cost whether the trucks are operated empty or full, and prices would depend primarily on the differences in demand on the two routes. If the combined revenues from the transport service between A and B and B and A is not sufficient to cover

the combined cost, the number of trucks (M) used to provide the service would be reduced and the prices in both directions would rise until joint costs are covered. Surplus revenue over joint costs would have the opposite effect. The most important point to grasp here is that there are differences in the demand for transport in each direction and that different prices should be charged in order to establish equilibrium. This means that knowledge of the price elasticity of demand and of the magnitude of the joint costs can solve problems that have arisen as a result of the untraceableness of costs (Button 1982:88–90).

In cases where joint costs are very heavy, marginal cost pricing in transport could be theoretically explained on the basis of figure 8.6. The joint marginal costing of the return journey from A to B and from B to A is represented by MC_1. For the sake of convenience we are assuming that the marginal costs have remained constant over the relevant output range. D_{ab} is the demand for transport between A and B, and D_{ba} is the demand between B and A. The economically efficient level of output is found at the point where price is equal to marginal costs. To determine the total effective demand SD, the two individual demand curves are added. The economically efficient level of output is at q_1, the point where the summed demand curve SD intersects the marginal costs curve MC_1 and at a price of p_1. In reality p_1 is the sum of the individual prices, namely p_{ab} for the trip from A to B, and p_{ba} for the trip from B to A. Because p_{ab} is greater than p_{ba}, a higher portion of the joint costs is borne by users between A and B than by users between B and A. This happens simply because the demand for transport between A and B is greater than the demand for transport between B and A. In equilibrium the prices in the two directions would not necessarily be the same. The greater the demand in a direction, the greater the proportion of the joint costs the users travelling in that direction would pay. In the extreme case the total joint costs would be borne by the users travelling in one direction. If MC_2 instead of MC_1 represents the joint costs, the economically efficient price and the output would be p_2 and q_2 respectively, in which case the total incremental cost of offering a return trip would be recovered from the users of the route from A to B.

When joint costs are present a competitive market would generate economically efficient prices and therefore lead to an economically efficient level of output. In equilibrium, the joint marginal cost is equivalent to the sum of the individual prices. If the sum of the prices exceeds the marginal costs, it would pay the enterprises to expand their services by increasing the total number of trips (new carriers could also enter the market), causing the summed price to fall. If the summed price is lower than the joint costs, the service would be curtailed, causing the summed price to rise (Bonsor 1984:23–26).

Although the prices generated by joint production are in proportion to the relative price elasticities of demand, this does not amount to price discrimination, which in fact involves different commodities that have nothing in common (on the demand side). Nevertheless the two are sometimes difficult to distinguish from one another, as the controversy between Taussig and Pigou at the beginning of the century illustrates. Taussig ascribed the large number of different rail tariffs to joint production, whereas Pigou regarded this as the result of price discrimination to cover joint costs.

The same arguments could be advanced with regard to airfares at the moment. According to Philips (1983:8), the difference in rates for passengers and freight can be ascribed to joint production. Because a passenger aircraft of necessity has empty cargo space available when it is conveying passengers (although the opposite is not true), air transport costs in relation to cargo are predominantly joint, and they are exclusively joint when it comes to luggage. A carrier is in no way obliged to offer its passengers a choice between first class and tourist class accommodation, and in this case Pigou is probably right and the airline's actions amount to product differentiation in order to apply price discrimination.

9.9 CONCLUSION

Pricing in transport is a complicated matter. Far from being fixed in an arbitrary manner, transport prices have to be very carefully calculated. In the transport sector it is also necessary to keep track of external costs. This question is addressed in the following study unit.

SELF EVALUATION

(1) Explain the meaning of the concept "economic efficiency" in welfare economics. (4)
(2) Discuss in full the economic reason for marginal cost pricing. (30)
(3) Should prices be based on long term or short term marginal cost? Discuss. (10)
(4) Is the choice of a unit of output of any importance in marginal pricing? Explain. (5)
(5) Discuss second-best pricing in full. Use the appropriate graphics to illustrate your discussion. Can second-best pricing be optimal in the economic sense? Explain your answer. (15)
(6) Define price discrimination. What conditions should be met before price discrimination can be applied? (12)

(7) Define and explain the various degrees of price discrimination. Do you think that any of these forms of price discrimination occurs in transport? Explain your answer. (9)

(8) How could a private monopolist increase his profits by applying price discrimination? Use the relevant graph to illustrate your answer. (10)

(9) Would a public monopolist's approach to price discrimination differ from that of a private monopolist? Explain your answer with reference to an enterprise with decreasing long-term unit costs, such as a railway enterprise. (15)

(10) Discuss pricing when a large proportion of the costs are joint. (15)

PART 3

Infrastructure investment in transport

OVERVIEW OF PART 3

Up to this point you have studied Transport Economics chiefly in the short term. In the long term there is additional investment and the capacity of the means of production are enlarged. In this study unit we shall be looking at investment in immovable infrastructure. We have again approached this discussion from the point of view of welfare economics.

STUDY UNIT 10

Infrastructure investment in transport

CONTENTS

OBJECTIVES

After studying this study unit you should be able to:

- explain the difference between the private and public approaches to investment with the aid of the relevant graphs
- explain the unique situation of investment in indivisible transport infrastructure with the aid of the relevant graph
- explain the theoretical aspects of cost-benefit analyses and the methods involved
- use graphical illustrations to explain how the benefits and costs of expanding transport capacity can be determined
- discuss national income as an alternative measure for evaluating investment

Market imperfections such as monopoly power and externalities, and the pervasive influence of transport, make it necessary to find suitable methods of evaluating transport investment. In this study unit we shall take a closer look at a few economic approaches.

10.1 INTRODUCTION

In the preceding study units special emphasis was placed on achieving economic efficiency through pricing according to marginal costs. Emphasis was also placed on the short-term aspects of efficiency, that is the optimal utilisation of existing facilities and services. In this study unit attention is given to the long-term aspects of economic efficiency, namely investment in new facilities or the expansion of existing facilities. In the case of road transport, airlines and shipping enterprises, which are mainly privately owned, investment decisions are normally taken on financial grounds. In most countries, however, road, rail and harbour infrastructure belong to the state, which follows a far broader approach to economic efficiency.

Transport enterprises operate in a market in which there is a considerable gap between private costs and benefits on the one hand and social costs and benefits on the other. Market imperfections, such as the existence of monopoly power and external costs, are reflected in market prices, and consequently the benefits and costs expressed in market prices are not accepted as an indication of the underlying social costs and benefits of investment. To enable the government to make the best possible investment decisions,

a framework is required that very clearly indicates the difference between social and private costs. Cost-benefit analyses provide a framework of this nature.

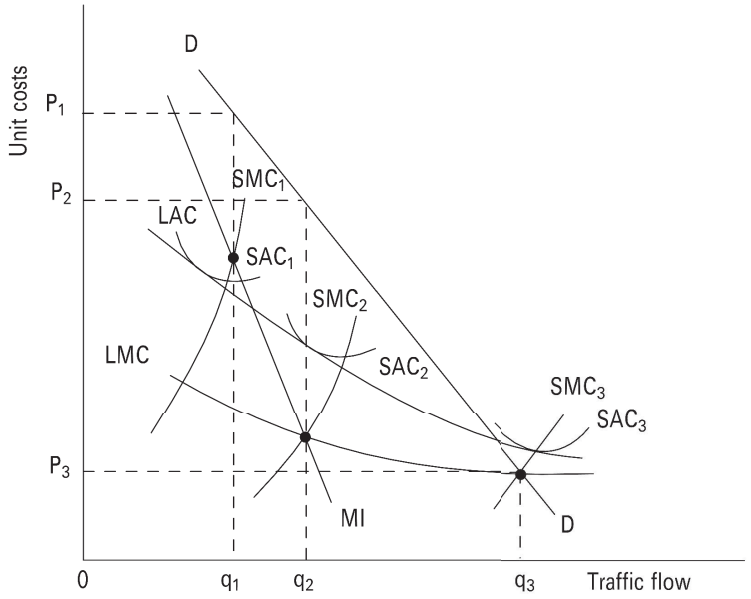
In this study unit, we deal mainly with the economic principles of investment, with special emphasis on transport infrastructure and the theory of cost-benefit analyses.

10.2 THE ECONOMIC PRINCIPLES OF INVESTMENT

Economic theory provides relatively straightforward guidelines for investment. Essentially, they involve pricing and output decisions under conditions where capacity is no longer a limiting factor. Figure 10.1, for example, illustrates the case of a profit-maximising airline with a short-term average cost curve SAC_1, and a short-term marginal cost curve SMC_1, confronted by a demand curve DD and a marginal income curve M_1. In order to maximise profits the airline has to fix its price at p_1 and offer q_1 seat kilometres. The long-term marginal costs (LMC) are, however, lower than marginal income at this point, and in view of its exceptional fleet size, this serves as an inducement to the airline to expand its capacity in the long term. Increased seat availability will force the price down, but the enterprise in our example (as you probably noticed) is experiencing declining long-term costs, which will possibly make it more economical to increase the size of the fleet. Figure 10.1 shows that the fleet could be expanded to correspond to SAC_2 and SMC_2 (the assumption throughout is that profit-maximising prices are charged). The long-term optimum is achieved when marginal income is equal to long-term and short-term marginal costs and profit is consequently also maximised. However, if the enterprise is interested in economic efficiency (maximum social benefit) rather than in commercial profit maximisation and prices were determined according to the marginal costs principle, the SAC_3 and SMC_3 curves become relevant because this objective requires that D = SAC = SMC. Greater capacity is needed for this and if the airline expands to SMC_3 and SAC_3 , the output would be higher at q_3 and the price lower at p_3. At this price the social surplus (consumer plus producer surplus) is maximised, notwithstanding the fact that a financial loss would be suffered because p_3 is lower than LAC.

FIGURE 10.1

Optimal investment



Source: Adapted from Button (1982:193).

Here you should distinguish clearly between a private monopolist and a public monopolist. The former pursues profit maximisation, whereas a public institution aims to achieve maximum social welfare. So although a loss may be suffered at p_3 it is economically efficient to allow the traffic flow to increase to q_3. At this point the full demand is being satisfied and the long-term marginal costs covered, although the long-term average costs would not be covered. If the supplier of the air transport service were a private

enterprise, it would have to insist on a state subsidy since the price it would be able to charge would not cover its long-term average costs.

Clearly where a public body is responsible for investment decisions in the transport sector, it would use broader economic criteria and look past the financial or market prices to the shadow prices. Shadow prices are more formally defined as the increase in welfare that results from any marginal change in the availability of factors of production. The process of shadow pricing therefore requires, firstly, a well-defined social welfare function in the form of a mathematical expression of a country's objectives, so that marginal change can be evaluated, and secondly a clear grasp of the limitations and policies that determine a country's development both now and in future, and therefore of the existing or projected circumstances under which marginal change will take place (Squire & Van der Tak 1975:49).

The source costs that reflect the scarcity value of factors of production are important in this approach. It may be possible to use unemployed labourers in the building of an airport, which could help to reduce unemployment; in South Africa this could also make a contribution to the RDP. The opportunity costs of the unemployed labour are effectively nil and are not taken as a cost in social cost-benefit analysis.

But a commercial or private enterprise would appreciate the input of the unemployed labourers at the wage it would have to pay. Another very important distinction between the commercial approach and the social cost-benefit analyses is that the latter recognise the distributional effect of investment by using weighted values.

Although the principles are simple, their application to transport causes a number of problems. Investments are frequently not divisible, and consequently the LAC and LMC curves are not as smooth as the figure shows, but are rather made up of a series of disjointed segments (or even points) that do not intersect the demand curve. This is a common situation in transport and one that poses many operational problems. It is not difficult, for instance, to envisage routes where the available vehicles (aircraft, buses etc) are either too small or too big for optimal efficiency. This is just as common in the case of infrastructure, where for example a two-lane road is unable to cope with normal demand but a three-lane road is too big. Another problem is deciding what "costs" should include. In figure 10.1 profit maximisation and marginal cost pricing are represented in the same diagram. In practice it would not be difficult because the public sector has to take a far wider range of costs into account when making investment decisions than do enterprises that are geared solely to profit. Furthermore, the diagrammatical analysis assumes, irrespective of any other criteria that apply, that prices are optimal in the short term and can consequently serve as guidelines for investment decision making. Another factor is that, despite the sophistication of prediction techniques, the supplier of transport often does not know for certain what the long-term demand curve with which he will be confronted, will look like. Because of the fluctuating nature of transport demand (especially the long-term cycles associated with national and international conditions), this curve is a stochastic (probability) rather than a deterministic phenomenon.

In view of all these difficulties, together with a general lack of adequate information among transport suppliers on the current or future scale of their costs, it is not surprising that the analysis of transport investment is receiving a lot of attention. The high cost and long-term implications of investment in transport facilities such as roads, railway lines and harbours have been extensively studied. At a theoretical level these issues pose numerous problems, because facilities are made available to users at prices that are not cost-related or are even made available free of charge. In addition there are widespread side effects for nonusers living in the surrounding areas.

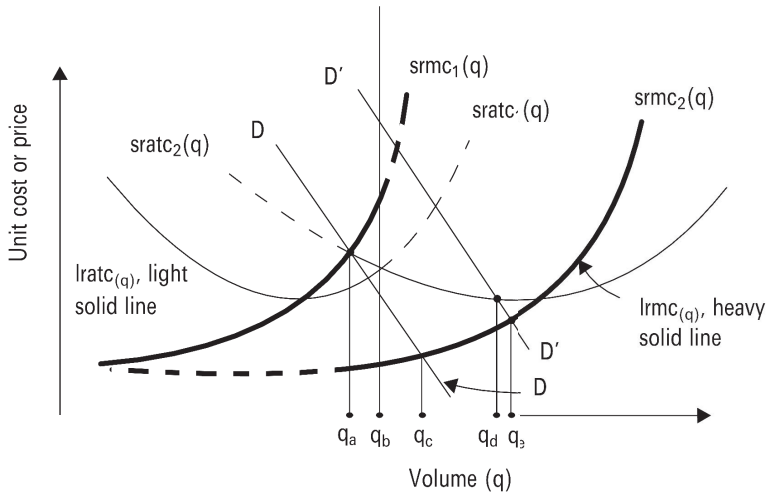
10.3 TRANSPORT INVESTMENT IN INDIVISIBLE INFRASTRUCTURE

In the previous section we mentioned the problems associated with the application of the economic principles of investment. In the case of transport infrastructure and particularly airports, harbours, roads and railway lines, infrastructure is highly indivisible. As you will remember from study unit 5, this means that a smooth or marginal investment cannot take place, but that it is usually necessary to invest in units of more than one. For example, one cannot build an extra metre of runway or road; it would have to be kilometres at a time. There are therefore no marginal costs in the normal sense of the word, but rather incremental costs, because expansion is carried out in increments and not in units. As a result these cost curves are jagged instead of being smooth. In figure 10.2 the "saw tooth" areas are represented by the thick dark lines.

Despite the "saw tooth" trend, the basic principles of investment remain the same. The facility has to be enlarged up to the point where the incremental social benefits of the expansion exceed or are equal to the long-term incremental costs. A further requirement is that the long-term net social benefits should not be negative. This requirement simply means that the expansion should not be detrimental in comparison to the situation before the expansion took place. It is not permissible to place society as a whole in a worse position than previously.

FIGURE 10.2

Transport investment with long-term incremental costs in indivisible infrastructure



With reference to figure 10.2 above, D'D' is the higher demand where marginal benefit is equal to demand. Investment is therefore justified up to point q_e, because here demand (and marginal benefit) are equal to the long-term marginal cost ($lrmc_q$). At this point this curve lrmcq is also equal to $srmc_2q$, which is the short-term marginal cost. The short-term and long-term marginal costs coincide in this way because we are now dealing with incremental costs that are discrete and not continuous. Because the marginal (or more correctly the incremental) costs have been covered by the marginal (or incremental) benefit – the latter is indicated by the demand curve – up to the point qe, there has been economic efficiency. The requirement marginal revenue = marginal cost has been met. (In this discussion marginal cost is used as a synonym for incremental cost, although you should know that when we are dealing with infrastructure incremental cost is the better term.)

Let us now examine demand curve DD. Here we are dealing with a lower demand. It is again assumed that DD also represents the marginal benefit. Clearly we need to expand the facility to the quantity q_a because DD = $lrmc_q$. This means that the marginal benefit is equal to the marginal cost. The big question is whether we should move from facility 1 to facility 2. Remember here that the long-term cost curve is not continuous in the case of indivisibility, but consists of two discrete curves that correspond partially to their respective short-term curves.

Therefore, we cannot move from quantity qa to qb, but instead have to "jump" to a point to the right of q_b. Let's take a look at point q_c. The answer is that we would only move from q_a to q_c if the incremental benefits are at least equal to or are greater than the incremental long-term costs. This means that the sum of all the incremental benefits must be greater than or equal to the sum of all the incremental long-term costs. We will then require more specific information on the relevant demand and cost functions.

It will be clear to you that with indivisible infrastructure the calculation of investment is far more complicated than in the case of divisible facilities, which we discussed in section 10.2 and figure 10.1.

10.4 THE THEORY OF COST-BENEFIT ANALYSES

Most of the consequences of changes in transport infrastructure extend over a wide area and a long period, and require refined methods of project evaluation and comprehensive decision-making techniques. The principles underlying cost-benefit analyses have already been discussed above but it should be remembered that these principles are based on a set of very complicated assumptions, which make their application far more difficult than it at first appears.

In the above outline we indicated that cost-benefit analyses are geared to the identification of projects that maximise social benefits rather than monetary returns. One of the thorniest problems here relates to the interpersonal comparisons of welfare. The question is whether it is really possible to contend that welfare has improved if one group has improved its position at the expense of another. In transport users commonly benefit at the expense of nonusers. In cost-benefit analyses we try to circumvent this problem by making use of hypothetical compensation tests. Because personal priorities can only be ranked ordinally, comparisons are,

strictly speaking, extremely limited. The Pareto criterion which underlies welfare economics states that an economic action can be said to be socially desirable only if at least one entity (individual) benefits without anyone else suffering a diminution of welfare.

In welfare economics, and in assessing investment in transport infrastructure, cost-benefit analyses are therefore used. This module does not aim to provide a comprehensive explanation and discussion of this technique. In a subsequent module, TRL302 (Transport planning and investment) you will learn about this in greater detail. Nevertheless, the following points deserve a brief mention.

Cost-benefit analyses are applied to mutually exclusive projects. This means that either one project or the other can be implemented. The technique is applied to infrastructure projects such as roads, airports etc. This technique is used to determine

(1) that no project of which the total net benefits are not positive can be regarded as economically acceptable
(2) that the project with the highest total net benefits should be selected as the most suitable. The determination of the relevant benefits and costs will be discussed in the following section. We should like to briefly mention the different types of cost-benefit methods.

10.4.1 Cost-benefit ratio

In this method a ratio is used to set off all the benefits against all the costs. If the ratio is over 1, the benefits of the project exceed the costs. The project is therefore economically acceptable.

10.4.2 Net current value method

All the benefits and costs of the projects are separately discounted to their current values. The project with the highest current value is regarded as the most suitable project.

10.4.3 Internal return method

According to this method a rate is calculated at which the benefits would equal the costs. This rate is then assessed to determine whether the project's rate of return is acceptable.

All these methods are open to criticism and they should be applied with great circumspection. Nevertheless they do supply a quantitative basis for very complex decisions. They can usually provide guidance to political or public decision makers who have to take final decisions on transport infrastructure.

10.5 DETERMINING THE RELEVANT BENEFITS AND COSTS

We have pointed out in several places in this study guide that any transport project, such as the extension of an airport, will generate a stream of direct and indirect benefits and costs. The provision of additional transport capacity, such as additional runways or terminal facilities at an airport with heavy congestion, leads to a decrease in congestion. In figure 10.3 the aggregate demand curve for all passengers using an airport is represented by D_1. Prior to any expansion of capacity, q_1 units are delivered at a price p_1, which is equivalent to the social marginal cost SMC_1. The social marginal cost consists of production and congestion costs, and the existing level of consumer surplus is represented by abp_1. An expansion of capacity will lead to a decrease in congestion, causing the marginal costs curve to move to SMC_2 and the demand curve to move from D_1 to D_2. The optimal price decreases from p_1 to p_2 and the quantity demanded increases from q_1 to q_2. The shift in the demand curve is due to an improvement in the quality of service, as a result of which consumers are willing to pay a higher price.

The total direct benefits of investment are represented by the area (cabd) + (dq_1q_2f). These direct benefits can be broken down into the following categories:

10.5.1 Benefits to existing users

Existing users receive a higher quality of service in view of the decrease in congestion. The net increase in existing users' surplus is equal to the area cabd. The general magnitude of cabd is approximately equal to the value of the time saved by existing users of the airport. It is assumed that these savings are employed productively.

10.5.2 Benefits of additional traffic

The decrease in congestion leads to an increase in the total traffic volume from q_1 to q_2. There are three distinct sources of additional traffic:

(1) Passengers who formerly used other airports may switch to using the new facilities. This traffic is known as derived traffic and it arises because the new facilities offer greater benefits than the ones formerly used.
(2) Passengers who previously used other means of transport may be attracted to air transport by the shorter travelling time.
(3) Existing marginal users will undertake more air transport. The total benefit enjoyed by additional users is equal to the area dq_1q_2f, and their social surplus or net benefit is equal to the area def.

Besides the direct benefits to users, there are a number of indirect benefits as well. Air carriers will enjoy a saving in the cost of factors of production as well as a result of the decrease in congestion. In general these would comprise savings on fuel, lower usage-linked maintenance costs and the improved utilisation of aircraft. Also, the operating costs of other facilities will decrease because much of the traffic will use the improved facility instead and consequently there will be a decrease in congestion at the other facilities.

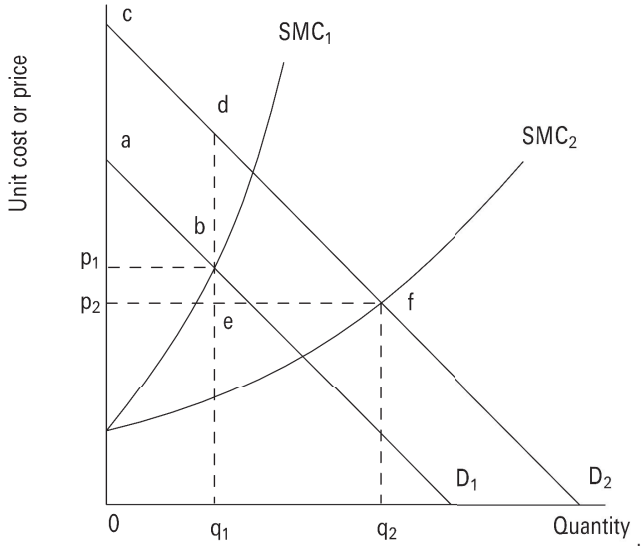
The cost of the expansion or improvement in the existing facility can similarly be divided into direct and indirect components. The direct cost of the project comprises mainly depreciation, financing and maintenance costs. The indirect costs include the higher noise and air pollution costs of the enlarged facility (Bonsor 1984:41–43). (Revise section 8.2 on the distinction between technological and monetary externalities.)

10.6 THE EFFECT OF TRANSPORT INVESTMENT ON NATIONAL INCOME

The effect of transport on national income may be used as an alternative criterion for evaluating investment. For the purposes of this discussion national income may be regarded as the sum of the market values of the goods and services purchased by business enterprises for investment purposes, by households for consumption and those purchased by government bodies. The change in national income could be used as a substitute for the combined consumer and producer surplus. Apart from the practical difficulties involved in estimating possible changes in national income in relation to transport investment, the fundamental question arises whether this approach really offers a reasonable and acceptable alternative guide to the relative desirability of alternative investments.

FIGURE 10.3

The benefits and cost of capacity expansion

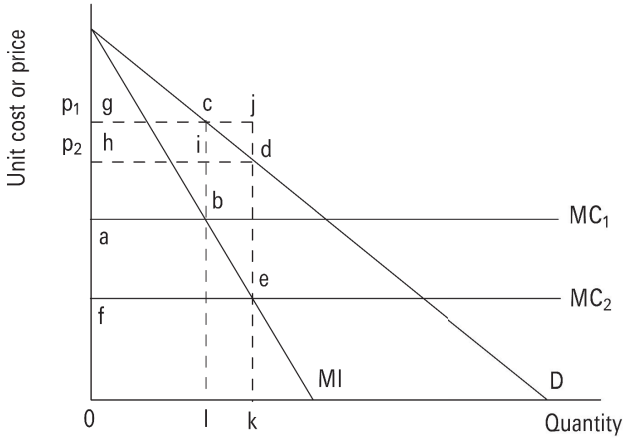


In figure 10.4 it is assumed that D represents the demand for a transport service and that D will not shift following an expansion in capacity. It is also assumed that the transport enterprise has monopoly power and is exercising it in its pricing policy. It is clear

from the figure that an expansion in capacity under these circumstances would reduce marginal costs from MC_1 to MC_2, and consequently the social surplus will increase by abcdef and the profit by (abidef) + (gcih). The reduction in costs will also cause the national income to rise. If the prices that prevailed before investment are used as a criterion, the increase in national income is equivalent to lcjk. If, however, the price after investment is used as a yardstick, the increase is only lidk. There is no reason why the social surplus criterion should be the same as the national income criterion, except under very unrealistic conditions. Nor will it necessarily correspond to the newly generated profit. What is of greater practical importance, however, is that there is no reason why alternative investment possibilities should be ranked consistently in the same order by the various methods.

FIGURE 10.4

The national income approachn



The reason why the results of the social surplus and national income approaches (and the financial criteria) need not correspond is that they all measure different things. Social or consumers' surplus includes leisure benefits as well as the diminishing marginal utility that accompanies an increase in the use of transport. The national income criterion includes neither; it concentrates exclusively on goods and services sold in the conventional markets and assumes either a fixed pre-investment price or a fixed post-investment price.

Mohring (1976) has demonstrated that the only time when the two approaches will produce the same answer is when, firstly, a change takes place that increases the output obtained from a given set of primary resources, secondly when the primary resources themselves do not change and thirdly the same pricing rules are used in calculating the consumer surplus and the national income. Patently, the national income measure is likely to differ in practice from both the financial and the social surplus measures of benefit and it is therefore simply an alternative – and not necessarily better – method of evaluation.

Given the practical difficulties experienced with estimates in most developed economies, their use is limited. In developing economies, however, estimating can offer a more viable method of appraisal if income distribution and welfare considerations are felt to be less important than boosting the national income. In such circumstances preference may be given to transport projects that help boost the national income (Button 1982:216–218).

SELF EVALUATION

(1) Why are market prices generally not a sound basis for evaluating transport investment? (5)

(2) Explain the basic economic principle that applies to investment by means of a graph that clearly indicates the relevant cost curves. Distinguish clearly between the level that investment will reach if the objective is profit maximisation and the level that it will reach if the purpose is maximum social welfare. (20)

(3) By means of a graph, explain the nature of investment in indivisible infrastructure and its attendant difficulties. (30)

(4) Investment in transport infrastructure is accompanied by benefits and costs. By means of a graph, explain how the relevant benefits of capacity expansion can be determined. Indicate what the direct and indirect costs are. (15)

(5) Does national income offer an alternative means of evaluating transport infrastructure? (15)

Baumol, WJ, Panzar, J & Willig, RD. 1982. *Contestable markets and the theory of industry structure.* New York: Harcourt Brace Jovanovich.

Bay Area Transit District, BART. 1968. *Catalyst for Bay Area Planning 10(2).*

Bell, G, Blackledge, DA & Bowen, PJ. 1983. *The economics and planning of transport.* London: Heinemann.

Bigham, TC & Roberts, MJ. 1952. *Transportation – principles and problems.* 2nd edition. New York: McGraw-Hill.

Bonsor, NC. 1984. *Transportation economics: theory and Canadian policy.* Toronto: Butterworths.

Bowersox, DJ, Clabro, PJ & Wagenheim, GD. 1981. *Introduction to transportation.* New York: McGraw-Hill.

Button, KJ. 1982. *Transportation economics.* London: Heineman.

Button, KJ & Gillingwater, D. 1986. *Future transport policy.* Beckenham, Kent: Croom Helm.

Clark, JM. 1940. Towards a concept of workable competition. *American Economic Review* 30(2):241–256.

Clark, JM. 1961. *Competition as a dynamic process.* Washington, DC: Brookings Institution.

Commercial Transport, February 1987:3.

Commercial Transport, November 1987:11.

Davies, JE. 1986. *Contestability and the Australian liner trades.* Occasional Paper 78, Bureau of Transport Economics. Canberra: Australian Government Publishing Service.

Dearing, CL & Owen, W. 1949. National transportation policy. Washington DC: The Brookings Institution.

Fair, LF & Williams, EW. 1981. *Transportation and logistics.* Revised edition. Plano, TX: Business Publications.

Faulks, RW. 1973. *Principles of transport.* London: Allen.

Glasson, J. 1978. *An introduction to regional planning.* London: Hutchinson.

Hicks, JR. 1940. The valuation of social income. *Economica* (7):105–124.

Isard, W. 1956. *Location and space economy.* New York: Wiley.

Isard, W. 1960. *Methods of regional analysis.* New York: Wiley.

Kahn, AE. 1971. *The economics of regulation: principles and institutions, vol. 2.* New York: Wiley.

Kaldor, N. 1939. Welfare proposition and interpersonal comparisons of utility. *Economic Journal* 49:549–552.

Kanafani, A. 1983. *Transportation demand analysis.* New York: McGraw-Hill.

Kaysen, D & Turner, DF. 1959. *Antitrust policy: an economic and legal analysis.* Cambridge, MA thesis: Harvard University Press.

Kneafsey, JS. 1975. *Transportation economic analysis.* Lexington, Mass: Lexington Books.

Kolsen, HM. 1968. *The economics and control of road rail competition.* Sydney: Sydney University Press.

Korver, W, Ruijgrok, CJ & Gwilliam, KM, 1992. *Kostentoerekening en prijsvorming bij vervoerbedrijven.* Leiden-Antwerpen: Stenferd Kroese.

Landon, CE. 1951. *Transportation: principles, practices and problems.* New York: Sloane.

Lefebre, L. 1958. *Production, transport and industrial location.* Amsterdam: North Holland Publishing.

Lichfield, N & Chapman, W. 1968. Cost-benefit analysis and road proposals for a shopping centre: a case study: Edgware. *Journal of Transport Economics and Policy* 2:280–320.

Lipsey, R & Lancaster, K. 1956. The general theory of the second best. *Review of Economic Studies* 26:11–32.

Little, IMD. 1950. *A critique of welfare economics.* Oxford: Oxford University Press.

Meyer, JR, Pech, MJ, Stenason, J & Zwick, C. 1959. *The economics of competition in the transportation industries.* Cambridge, Mass: Harvard University Press.

Milne, AM & Laight, JC. 1963. *The economics of inland transport.* 2nd edition. London: Pitman.

Mishan, EJ. 1967. Interpretation of the benefits of private transport. *Journal of Transport Economics and Policy* 1:184–189.

Mohring, H. 1976. *Transportation economics.* Cambridge, Mass: Ballinger.

Owen, W. 1964. *Strategy for mobility.* Washington: Brookings Institution.

Panzar, JC & Willig, RD. 1981. Economies of scope. *American Economic Review* 71(2):268–272.

Pegrum, DF. 1973. *Transportation: economics and public policy.* Homewood, Ill: Irwin.

Philips, L. 1983. *The economics of price discrimination.* Cambridge, Mass: Cambridge University Press.

Pigou, A. 1920. *The economics of welfare.* London: Macmillan.

Ramsey, FP. 1927. A contribution to the theory of taxation. *Economic Journal* 37 (March):47–61.

Rothenberg, J. 1970. The economics of congestion and pollution: an integrated view. *American Economic Review, Papers and Proceedings* 60:114–121.

Ruppenthal, KM & McKinnel, HA. 1966. *Transportation and tomorrow.* Stanford: Stanford University Press.

Sampson, RJ, Farris, MJ & Shrock, DL. 1985. *Domestic transportation: practice, theory and policy. 5th edition.* Boston: Houghton Mifflin.

Schumer, LA. 1967. *Elements of transport.* 3rd edition. Sydney: Butterworths.

Shahia, M. 1980. Stedelike vervoer en verkeersbeleid met besondere verwysing na die invloed van parkering daarop. MCom-verhandeling, Universiteit van Suid-Afrika, Pretoria.

Sharkey, WW. 1982. *The theory of natural monopoly.* Cambridge, Mass: Cambridge University Press.

Shepherd, WJ. 1984. *Contestability vs. competition.* American Economic Review 74(4):572–587.

Smith, DM. 1966. A theoretical framework for geographical studies of industrial location. *Economic Geography* 42(2).

Smuts, C. 1989. 'n Vervoerekonomiese ontleding van die inhoud en implikasies van die nasionale vervoerbeleidsdoelwitte vir staatsgesubsidieerde busvervoer. DCom-proefskrif, Universiteit van Suid-Afrika, Pretoria.

Squire, L & Van der Tak, HG. 1975. *Economic analysis of projects. A World Bank Research Paper.* Baltimore, MD: John Hopkins University Press.

Stubbs, PC, Tyson, WJ & Dalvi, MQ. 1980. *Transport economics.* London: Allen & Unwin.

Suid-Afrika (Republiek). Departement van Vervoer. 1986. *Witskrif oor nasionale vervoerbeleid.* Pretoria: Staatsdrukker.

Taplin, JHE. 1983. Regulation, deregulation and the sustainability of transport monopolies. *The Logistics and Transportation Review* 19(1):31–44.

Tye, WB. 1987. The contestable market defense in freight antitrust cases. *Transportation Practitioners' Journal* 54(2):177–198.

Wahl, M & Hendrickson, G. 1984. *Transportation investment and pricing principles.* New York: Wiley.

Winston, C. 1985 Conceptual developments in the Economics of Transportation *Journal of Economic Literature* 23:57–94.