FACULTY OF ENGINEERING
DEPARTMENT OF CIVIL ENGINEERING
SECTION TRAFFIC AND INFRASTRUCTURE
KASTEELPARK ARENBERG 40, B-3001 HEVERLEE, BELGIUM

KATHOLIEKE
UNIVERSITEIT
LEUVEN

Course H 111 Verkeerskunde Basis

Basics of Transport Economics

February 2004
(Updated August 2007)

Prof. ir. L.H. Immers
ir. J.E. Stada

Preface

This text is part of the subject matter of the course on Basics of Traffic Engineering (H111) taught to the students in the Department of Civil Engineering at the Katholieke Universiteit of Leuven (Belgium).

The text is intended to introduce the student to some elementary principles of transport economics and may serve as a stepping stone to more advanced courses.

This is a preliminary version of the text. Comments and suggestions for improvement will be gratefully received.

Heverlee, February 2004

L.H. Immers
J.E. Stada

email:
traffic@bwk.kuleuven.ac.be

tel:
0032-16-321669
0032-16-321670

address:
Katholieke Universiteit Leuven
Departement Burgerlijke Bouwkunde
Sectie Verkeer en Infrastructuur
Kasteelpark Arenberg 40
B-3001 Heverlee (Belgium)

Translated from the Dutch by
Leni Hurley and Jim Stada (January 2005)

Contents

1. Introduction

For the economic and social functioning of society an adequate transport system is necessary. But it must be remembered that the transport sector produces a number of undesirable side effects including environmental problems, noise pollution, traffic accidents and congestion. On the one hand the transport sector offers economic and social advantages, on the other there are cost inducing disadvantages. When assessing the optimal size of the transport system, these advantages and disadvantages must be weighed against each other.

To meet the problems associated with increasing mobility, governments intervene with a number of policy measures[1$]. Examples are the modification of travel demand through spatial planning measures, modifying the modal choice through improved public transport, increasing the efficiency of transport and traffic and stimulating the technological development of vehicle innovation.

This chapter examines economically slanted policy measures that are designed to influence the way in which people look at mobility problems. We begin with the assumption that traffic causes social costs (such as congestion- and environmental costs) that are not or insufficiently levied on the user. These harmful effects, of which the consumer, due to the current transport pricing, is not or insufficiently aware, are called negative externalities. Some measures, road pricing for example, aim to internalise these negative externalities. That is to say: these effects become included in the overall price that the transport consumer must take into consideration when deciding whether or not to travel, or which mode of transport to use. As will be explained in this chapter, the social costs of a certain proportion of the traffic on our highways exceed the social benefits. There are a number of ways to remedy this undesirable situation. The most efficient way is by means of correct pricing.

As we said above, most governments already intervene by means of a number of policy measures. This is necessary and useful but most current measures still fail to rectify the difference between social costs and social benefits in parts of the transport sector. To rectify this situation, some form of road pricing should be added to all the various forms and sets of measures.

It must be born in mind that we need to consider the traffic-induced social costs both when optimising the use of the existing infrastructure and when planning new infrastructure. This chapter therefore also includes a short introduction to the field of social cost-benefit analysis.

[$] Notes refer to the references section at the end of the document.

The chapter contains three sections:

- We begin with a brief restatement of a number of basic fundamental ideas from microeconomics, with references to applications in the transportation sector. Special attention will be given to a particular branch of microeconomics, namely welfare economics.

- Next, we examine pricing in the transportation sector. With regard to pricing, governments aim to achieve an optimal *use* of *existing infrastructure*, in terms of maximising social welfare. In this chapter we show that this aim will be achieved when the additional (or marginal) benefits of additional use equal the additional (or marginal) costs inherent in this additional use. These costs need to be charged correctly and this implies, in many cases, a form of toll collection.

  We must emphasise that this toll charge is not intended to recoup past investments: users have already paid for this infrastructure via general taxes. The charge aims to persuade users to reconsider a trip if the costs of this trip exceed the benefits.

  This chapter does not deal with privately exploited toll roads. Companies with privately owned capital construct these roads. The tolls in this case aim to gain favourable returns on these private investments. This private interest does not necessarily coincide with the common interest of society at large.

- The last part of this chapter examines investment analysis. This does not deal with the use of existing infrastructure but with the decision to *construct new infrastructure*. There are a number of ways in which to explore the justification for new infrastructure. We give a short outline of the method of social cost-benefit analysis. The construction of new infrastructure is only justified when the total benefits, calculated over a certain period of usage, exceed the total costs of a project over the same period of time.

This basic course in transport economics is confined to some elementary principles of the application of economic science to the field of transportation. For more detailed information we refer to specialised courses in microeconomics and transport economics.

2. Basics of microeconomics

2.1 Efficiency and equity

Evaluating economic policy involves on the one hand assessing the efficiency of the policy measure and on the other hand assessing its fairness or equity. Efficiency in this context means that society as a whole reaps maximal benefits at given costs; equity means the 'fair' distribution of these benefits across the entire population. In other words: when we speak about efficiency we mean maximising the total size of the economic pie, equity deals with the fair portioning of the pie.

Equity or fairness is a highly subjective concept. There is no objective way in which to judge the fairness of a particular policy. This is why economic theory barely touches on this subject. The concept belongs primarily to the field of political philosophy.

But economic theory has a lot to say about the concept of efficiency. Criteria are needed to judge the efficiency of an economic measure. The most widely used criterion is that of Pareto. The *strict Pareto-criterion* states that there is an improvement in efficiency if a measure improves the welfare of at least one individual while no one looses out on the deal at the same time. A state of Pareto-efficiency applies when Pareto-improvements are no longer possible. Remember, however, that Pareto-efficiency says nothing about the division of prosperity amongst individuals. Both parity of incomes across the whole population and a very unequal division of incomes can be Pareto-efficient.

In practice, the application of the strict Pareto-criterion leads to problems. In every day life, economic measures always have winners and losers. To get around this problem, the economists Hicks and Kaldor proposed an alternative criterion namely the *potential Pareto-criterion*. According to this criterion economic efficiency is increased when the measure-introduced leads to a situation whereby the winners could *in principle* fully compensate the losers and still retain a net advantage. The application of the potential Pareto-criterion does not require or demand that actual compensation take place. The potential Pareto-criterion is central to the pricing measures and the cost-benefit analysis that are dealt with in this text.

2.2 Demand function: marginal benefits

Economic science uses the word "goods" in its broadest sense. Everything that helps satisfy a human need falls under this banner. In this sense a service provider, such as a transport company, produces goods. A transport provider offers his services on the transport market and consumers can decide if they want to avail of them or not. Producers and consumers do not always act as separate entities: sometimes they are one and the same person. An example is the driver of a motor car who at the same time produces and consumes the trip he makes in his own car.

People buy certain goods because they derive a certain utility from the consumption of these goods. People recognise that these goods represent a certain value to them and this reveals itself in their willingness to pay for the goods. We call this value the *benefit* of the goods.

We need a unit in which to express the benefit of goods. A monetary unit is the obvious choice because it can compare the degree to which various goods deliver various benefits.

Total benefit TB usually increases with increasing consumption. If q represents the amount of goods acquired, $TB(q)$ is a monotonically increasing function of q. But this function does not increase linearly with q. Experience shows that the additional benefit $\Delta(TB)$ derived from the acquisition of additional goods $\Delta(q)$ usually decreases for increasing q. In other words: the price $\Delta(TB)/\Delta(q)$ that people are prepared to pay for each additional unit of the goods decreases. This empirical finding is known as the first law of Gossen, a 19th century economist.

The additional benefit gained in the acquisition of an extra unit of goods is called the marginal benefit MB. The function $MB(q)$ is, therefore, a decreasing function.

The function $TB(q)$ is a discrete function, only defined for discrete values of q. We approximate it, however, by a continuous, differentiable function. The derivative of this $TB(q)$ with respect to q then becomes an approximation for $MB(q)$:

$$MB(q) = \frac{d(TB(q))}{dq}$$

Assume the cost of a particular good to be p, how much should we acquire? As long as the additional benefit of an extra unit acquired, i.e. the marginal benefit, exceeds the price p it is profitable to acquire additional units. If the marginal benefit decreases below p acquiring an additional unit leads to a loss. At price p, the rational consumer will, therefore, consume a quantity q of the goods such that $MB(q) = p$.

In economics we use the *individual demand function* to indicate the quantity of goods a person will acquire at a particular price. The above shows that the *individual demand function is identical to the function of marginal benefits MB(q)* for that person. For a particular q, the demand function therefore indicates the additional or marginal benefit received in the acquisition of an additional unit.

Alongside the individual demand function we also distinguish the *collective demand function* that results from horizontal addition (across the quantities) of all the individual demand functions. Analogously we can also speak of the collective marginal benefit and the collective total benefit. In future we will omit the prefix 'collective' and assume that we always refer to the collective population of all consumers.

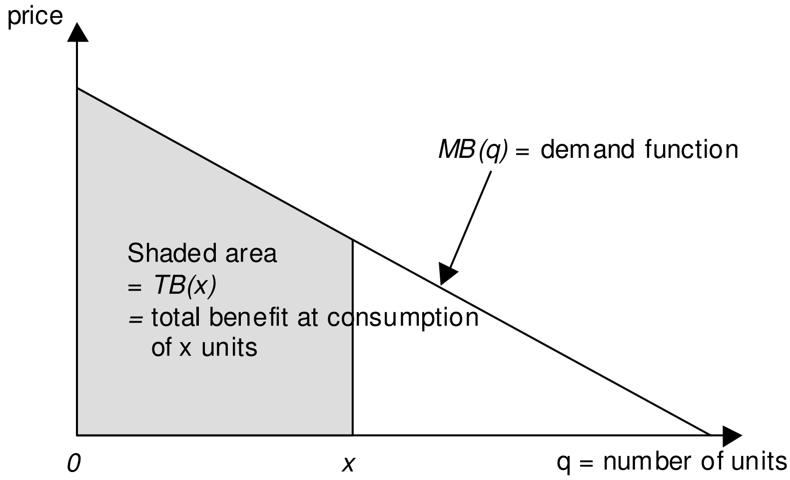

Figure 1 Demand function and determination of the total benefits

Since the demand function *MB(q)* is found by taking the derivative of the total benefit function *TB(q)*, the opposite also applies, namely that the total benefit can be found by integrating the demand function. This means that the area below the demand function from zero to a certain value of *x* represents the total benefit or the total value enjoyed by consumers when consuming that quantity[$] (see Figure 1[#]).

Determinants of demand

In the above we confined ourselves to the influence of price on the demanded quantity. We assumed that other determinants of demand remained the same. Price, however, represents only one of a number of determinants of demand.

A change in price alone implies a shift along the demand curve. If determinants other than price change, there will be a different demand curve: the whole demand curve shifts to another position.

The other determinants of demand:

- *Consumer income.* The general trend is for consumption of all kinds of goods to rise as available income increases. When this is the case, we speak of *normal goods*. There are exceptions where the consumption of certain goods decreases as income increases. We call these goods *inferior* goods. One example of an inferior good is that of bus transport. It has often been noted that the demand for bus transport decreases with rising income levels.

[$] In fact, we should use the area beneath a demand function that has been "compensated" for income effects. The difference between the ordinary and the compensated demand functions tends, however, to be negligable. For a more detailed discussion we refer to advanced courses on microeconomics.
[#] To keep things simple, we drew the demand function in Figure 1 as a straight line. The function generally decreases monotonically with increasing *q*, but it does not necessarily have to be linear.

- *The prices of related goods.* We distinguish substitute goods, complementary goods and independent goods. *Substitutes* are goods that fulfil a similar function to the good under consideration. Take, for example, the demand for car transport. A substitute for car transport is transport by rail. If the price of a train ticket were to increase, the demand for transport by rail would decrease (a shift along the demand curve) and the demand for car transport would increase (a shift of the entire demand curve). Overall, the demand curve for car transport will have risen. If an increase in the demand for one good also leads to an increase in demand for another good they are said to be *complimentary.* Cars and petrol, for example, are complimentary goods. When goods are not or barely related, we speak of *independent* goods.

- *Consumer preferences.* Advertising, for example, can alter consumer preferences. The media can influence opinions regarding the desirability of consumption of particular goods, including transportation. This, again, results in a shift of the entire demand curve.

2.3 Supply function: marginal costs

Producing and supplying goods and services requires the input of resources such as raw materials, time, space and energy. These inputs are lost to other applications. The production costs are now defined as the value of the best possible alternative use of these resources. For the sake of simplicity, the production costs are expressed in monetary units, as happened in the case of the consumer benefits.

Consider an individual producer of a particular good. As the production of goods increases, so do the *total (cumulative) costs TK*. If q represents the quantity of produced goods, this means that $TK(q)$ is an increasing function of q.

Total costs consist partly of fixed and partly of variable costs. Characteristic of fixed costs is their constancy: they do not depend on the quantity of goods produced. They represent the costs of capital goods such as machines, buildings, infrastructure, etc. that have to be paid for even if nothing was produced. Variable costs, on the other hand, represent the costs of variable production factors such as raw materials, labour, etc. that increase as production levels rise. Total costs TK represent the sum of the fixed costs FK and the variable costs VK:

$$TK(q) = FK + VK(q)$$

Economic theory distinguishes between short run analysis and long run analysis. Characteristic of short run analysis is the assumption that the amount of capital goods remains stable. The available production capacity is, therefore, also constant in the short term. Short run production can only be varied by changing the quantity of variable production factors. In long run analyses all production factors, including the amount of capital goods, are assumed to be variable.

Short run costs

The upper panel of Figure 2 gives the total costs to a producer as a function of the production volume *q*. Note the characteristic progression of the variable costs. The shape of the variable costs curve follows from the rule of "diminishing marginal product". Initially, before diminishing marginal productivity sets in, the variable costs rise at a rate less than or proportional to the volume of the variable production factors employed. At a certain stage, however, the diminishing marginal productivity begins to play a role and the variable costs rise sharply. The volume of capital goods is, after all, constant. If these goods are already optimally used, people in an industrial production environment start to get in one another's way. In traffic, from a certain level of flow cars begin to be in one another's way.

In determining supply, two pricing concepts play an important role, namely the average costs and the marginal costs.

The *average costs GK* equal total costs divided by volume of production. They are, therefore, the average costs per unit produced.

$$GK(q) = \frac{TK(q)}{q}$$

The average costs are shown in the lower panel of Figure 2. The average costs are easily derived from the graph of the total costs *TK*, shown in the upper panel. The average costs at a certain point equal the slope of the line starting at the origin to the point in question on the *TK*-graph. The average costs in, for example, point *E* of the *TK*-graph are measured by the slope of the line gk_E from the origin to point *E*. The graph for the average costs has a characteristic U-shape. Average costs for small *q* are high because small volumes still require relatively large fixed costs. The average costs achieve a minimum at q_{eff}, then rise again. Since the average costs are minimal at production volume q_{eff} this volume is called the *efficient scale of production* at the given investment in production capacity or, in other words, at the given level of fixed costs. The efficient level of production depends on the technology used in a specific trading sector.

The *marginal costs MK* are defined as the *additional costs* for the production of one extra unit of the good. For a continuous function *TK(q)* the difference quotient can be replaced by the differential quotient:

$$MK(q) = \frac{\Delta TK(q)}{\Delta q} = \frac{d(TK(q))}{dq}$$

The marginal cost curve is also shown in the lower panel of Figure 2. The marginal costs equal the slope of the tangent at the *TK*-curve. At point *F* in the graph, for example, they coincide with the slope of the tangent mk_F. The marginal cost function initially decreases or remains constant but it increases from a specific point, due to the rule of diminishing marginal product as alluded to above.

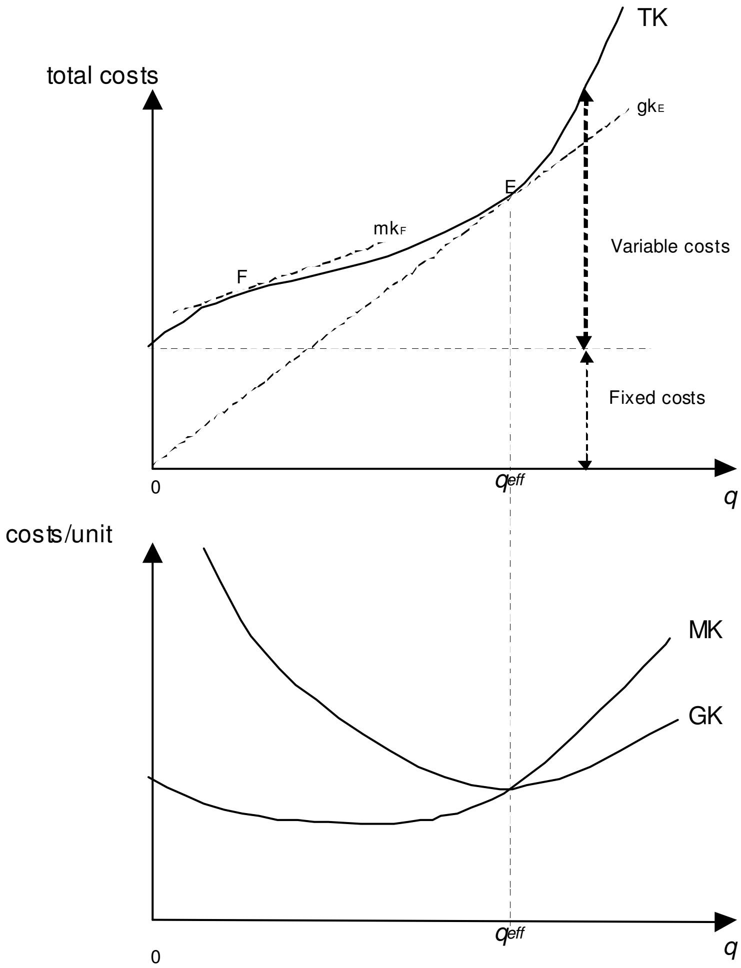

Figure 2 Short run costs

Figure 2 illustrates that at decreasing average costs the marginal costs curve lies below the average cost curve. But when average costs increase, the marginal costs curve lies above the average costs curve. This is no accident. It always applies when dealing with average and marginal quantities. Compare, for example, average costs with average scores across a number of exams. If the score of the next exam (the marginal score) lies below the current average, the new average will decrease. If the marginal score is higher than the current average, then the new average will also rise. This leads to the following important conclusion: *the marginal costs curve crosses the average costs curve at the minimum of the average costs curve, that is: at the point of the efficient scale of production* q_{eff}.

Now assume that the market price for a specific good equals *p*. Also assume that one particular producer cannot influence the market price and that he can sell his entire production at the market price. This happens when the producer competes with a large number of suppliers of the same product. Such a market is called a perfectly competitive market.

What volume of goods will the producer offer at the market price? As long as market price *p* exceeds the marginal costs incurred by the producer to manufacture an additional unit he will benefit by producing additional units. Once the marginal costs exceed price *p*, the producer looses on those additional units. At price *p*, therefore, the rational producer will offer a volume *q* of the goods for which $MK(q)= p$ applies.

In economics the so-called *supply function* is used that indicates what quantity of goods will be produced and offered at a particular price. The above shows that the *supply function is identical to the marginal costs function*. Thus, for a specific *q* the supply function gives the additional or marginal costs for the production of an additional unit of the good.

For a commercial producer, however, not the entire marginal costs curve does qualify as a supply curve. If the market price lies below the average costs, companies will make a loss. In the long run such companies are forced to discontinue production. For a commercial company, therefore, the supply curve is that part of the marginal cost curve that lies above the average cost curve, in other words, to the right of q_{eff}.

There also are economic sectors that operate in the production region to the left of q_{eff}. These are companies that have to undertake very high initial investment costs, so that the average costs continue to decrease over a large range and where relevant demand is confined to this range. Many (non-commercial) public utility companies belong to this category. We will return to this subject when discussing "natural monopolies" in section 2.7.

Since the supply function *MK(q)* is the derivative of the total cost function *(TC)*, the reverse applies in that the total costs can be found by integrating the supply function. This means that the area beneath the supply function, from 0 to a certain volume *x*, represents, except for an integration constant, the total costs involved in the production of the goods. The integration constant represents the total fixed costs, costs that do not vary with the number of units produced. In other words: the area beneath the marginal costs function from 0 to *x* equals the *total variable* costs for the production of *x* units. (see Figure 3)[$].

Until now we have examined the supply curve for a single company. The collective supply, or the short run *market supply*, is found by horizontal addition of all supply curves of the individual suppliers of the goods in question.

[$] The supply function in Figure 3 is drawn as a straight line. The real shape can, of course, differ, as explained in the text.

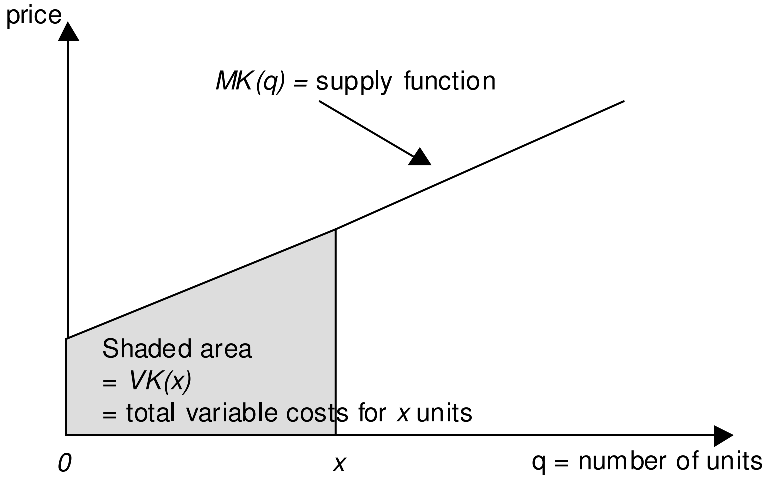

Figure 3 Supply function and determination of total variable costs

Long run costs

Until now we confined ourselves to the discussion of short run costs. When we look at the long run, both the quantity of variable production factors and the quantity of capital goods can change. In industrial production this means, for example, investment in additional machines. In road infrastructure it may mean an increase in the number of traffic lanes. Each quantity of capital goods has its own particular production capacity with its own costs structure and its own cost curves.

A manufacturer must decide on the scale of production of his firm, i.e. he must decide on the volume of his capital goods supply or production capacity. In infrastructure this means determining the optimal number of traffic lanes. If the average cost level decreases with increased scale we speak of *economies of scale*. Similarly we can have *diseconomies of scale* or *constant returns to scale.* Competition, together with the entry of new companies into the market or the cessation of others, leads to a long run tendency whereby each company grows to a size characteristic for a certain sector, a size where average costs are as low as possible.

Determinants of supply

It is not only the market price that influences supply. Other factors also determine the quantity supplied at a certain price. In general, a change in the prices of inputs to the production process may lead to a realignment of the entire supply curve. Changes in production technology may have the same effect. (Input prices are the prices of raw materials, labour and other elements used in production; in the transport sector we could for example think of personnel costs)

2.4 Specification of transport markets

The demand curve applies to one specific market that trades in a more or less *homogeneous* product. When one looks at the transportation sector, this fact must be born in mind, for this market, in a general sense, hardly ever deals in homogeneous products. If we want to apply the principles of microeconomics to transport we will need to clearly specify the kind of market we are talking about.

We could, for example, look at the market of home-based work trips by car between Brussels and Leuven. Compare this with the market for leisure time traffic between Brussels and the seacoast, which is an entirely different market with its own demand curve.

In general, we need to specify the following characteristics when defining a particular transport market:

- the transport relation,
- the transport mode,
- the trip purpose and
- the time at which the trip occurs (peak or off-peak, weekday, weekend).

But because classification into the categories above leads to a large number of different market segments, it is usual to apply an aggregate step at different levels, whereby the demand is aggregated over the factor that was left out of consideration. This also has implications for the unit in which the quantity demanded is expressed. The demand for car trips between Leuven and Brussels can be expressed in trip numbers. If the transport relation is no longer specified, the number of vehicle kilometres becomes a more appropriate variable.

2.5 Elasticity

The observed transport demand at a given moment is but one point on the demand curve. This observed demand is called the *manifest demand* for transport at that specific moment.

It is of practical importance to have quantitative information, not only about the manifest demand, but also about the shape of the demand function across a specific range. In other words: we want to know how the demand function changes if one of the determinants of demand changes. A transport operator, for example, will want to know how consumers react to price changes. If an increase in the price of a train ticket hardly influences the number of passengers, revenue for the rail company will increase. Price increases, on the other hand, could reduce the number of travellers so much that total revenue decreases. What should a railway company do if it seeks to increase its revenue? To answer such questions we can apply the concept of *elasticity*.

Elasticity measures the percentage change in a dependent (or response) variable that is due to a percentage change in an independent (or stimulus) variable. It is assumed that all other independent variables remain constant.

The concept of elasticity is broadly defined. We can use elasticity for the description of any kind of functional relation. Generally though, it is used in reference to demand- and supply functions.

Demand elasticity's

Demand elasticities are determined by observing how consumers react to a change in one of the determinants of demand (see page 8), where the other determinants are assumed to remain constant. One can simply ask consumers what they would do in a given situation. This is called *stated preference.* More reliable details are acquired when the actual reaction in a given situation is observed. This is called *revealed preference*.

The *price elasticity of demand* (or simply price elasticity) for a product shows the relation between the percentage change in the demanded quantity and the percentage change in the price:

$$price\ elasticity = \frac{\%\ change\ in\ demand}{\%\ change\ in\ price}$$

The price elasticity at a particular point of the demand function, therefore, gives an indication of the shape of the demand function in the vicinity of this point.

Example: When a price increase of 10% at a particular point on the demand curve results in a 5% decrease in the demanded quantity, the price elasticity in that point equals -5% / 10% = -0.5

In most cases the price elasticity of demand is negative. This is because the demand function is almost always a decreasing function. At a price elasticity of -1, price increase and decrease of the demanded quantity actually balance one another. At price elasticity values below -1, a change in prices has a relative large impact on the quantity demanded. In that case we speak of an *elastic* demand. When the price elasticity lies between 0 and -1, the demand is *inelastic*.

Total revenue, or the amount paid by consumers and received by producers, equals price multiplied by quantity. If demand is inelastic the demanded quantity is relatively unresponsive to price changes. A price rise will, therefore, increase total revenue; a lowering of price will decrease total revenue. If demand is elastic, the effect is exactly the opposite.

Price elasticity for various products can differ considerably. The demand for primary needs, for example, is quite inelastic. They are needed, no matter what the price is. Elastic demand curves on the other hand, are characteristic for luxury articles, they are more price-sensitive. In general, the demand for a product becomes more elastic, as:

- there are more alternative, replacement products,
- a larger share of income is spent on the product,
- people have more time to adjust to price changes

In transport we find, for example, that the demand for home-based work trips is much less elastic than the travel demand for leisure and shopping. This is logical for there are fewer alternatives to the consumer for his home-work trips than there are for other travel purposes.

The time horizon across which elasticity is measured is *very important* in transport. In general, the demand for transport in the short run appears to be relatively inelastic. In the long run, a change in transport prices can have a considerable influence on the demand and may even change the location of home and work.

Apart from price elasticity, other types of elasticity are defined in order to measure the influence of the other determinants on demand.

The *income elasticity of demand* shows the influence of a change in income:

$$income\ elasticity = \frac{\%\ change\ in\ demand}{\%\ change\ in\ income}$$

Normal goods show a positive income elasticity, inferior goods a negative income elasticity.

The cross price elasticity of demand is used to determine the effect of changes in the cost of related goods. Take a product *a*. We want to determine the change in demand for good *a* when the price of a related product *b* changes. The cross-price elasticity of *b* is then defined as follows:

$$cross\ price\ elasticity\ of\ price\ of\ b\ on\ demand\ for\ a = \frac{\%\ change\ demand\ for\ a}{\%\ change\ price\ of\ b}$$

When two goods are substitutes the cross-price elasticity is positive (example: if train fares increase, the demand for car transport increases also). It is negative for complementary goods (if petrol prices rise, the demand for car transport decreases). For non-related goods cross price elasticity equals zero.

<u>Supply elasticities</u>

For completeness, we mention that the concept of elasticity can also be used to describe the supply side. The definition of elasticity of supply is analogous to the elasticity of demand. Thus, the price elasticity of supply, for example, is defined as the percentage increase of the quantity supplied divided by the percentage increase in price. Supply elasticities are not used very often.

Examples of elasticity in transport

Elasticity values reported in the literature show large variations, depending on the degree of transport-market segmentation and aggregation level.

Table 1 gives an impression of a number of mean elasticity values that are used in the METS[2] model, a simulation model that describes the demand and supply of transport in London. The price elasticities shown in the table refer to the *generalised price* per trip. The generalised price not only includes the private car costs or the price for a ticket, but also the time costs converted into money prices. This is important since the literature often mentions elasticities that are based solely on the direct car costs or fare prices. Depending on the application, these elasticities reported in the literature, therefore, need to be adjusted because the behaviour of travellers is determined by the total costs of transport.

The diagonal from upper left to lower right shows the normal price elasticity. For example: the price elasticity for car transport is -0.30. This means that if the generalised price rises by 10%, then the demand for this type of transport will decrease by 3%.

Table 1 Examples of elasticity used in the METS model (London)

	Stimulus (generalised price)		
Response	Car price	Bus price	Tube price
Car demand	-0.30	0.09	0.057
Bus demand	0.17	-0.64	0.13
Tube demand	0.056	0.20	-0.50

The cross-price elasticities are shown outside the diagonal. These are positive, which means that the various transport options in the table are substitutes for one another. For example: the cross-price elasticity of the generalised bus fare on the demand for car travel is 0.09. This means that if the generalised price of bus transport rises by 10 %, the demand for car transport will increase by 0.9 %.

2.6 Demand and supply equilibrium in a perfectly competitive market

In Figure 4 the demand curve (the marginal benefits) and the supply curve (marginal costs) are brought together. The point of intersection *S* of both curves determines the market equilibrium.

This equilibrium is the only stable combination of price and quantity in a perfectly competitive market. A perfectly competitive market (sometimes called a competitive market or a free market) is characterised by the presence of a very large number of suppliers of a homogeneous product. A single producer is, therefore, unable to influence the market price.

When the market price exceeds the equilibrium price *p* we get a market surplus of goods. To avoid being left with surplus goods, producers will react by lowering their prices. If the market price is below the equilibrium price, this leads to a shortage.

Producers will react to this situation by increasing their price. In both cases there will be a movement towards the stable market equilibrium.

As explained before, the area beneath the demand curve, from the origin to a particular quantity *x*, equals the total benefit to the consumers from the consumption of quantity *x*. The amount paid by the consumer for this quantity equals the quantity *x* multiplied by the price per unit *p*. The total amount paid by all consumers together, is, therefore, given by the surface of the rectangle *opSx* beneath the price line.

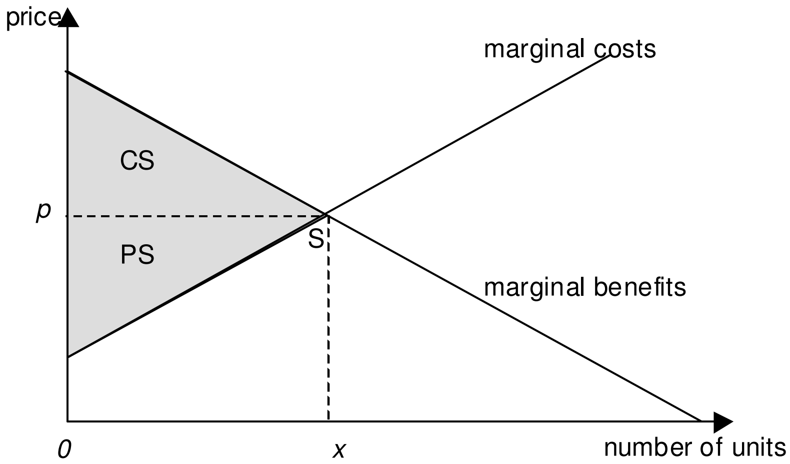

Figure 4 Consumer surplus and producer surplus

The difference between total benefits and the amount paid, in other words, the difference in both surfaces mentioned above and indicated in Figure 4 by CS, is a measure for the joint advantage for all consumers, the *consumer surplus*. The consumer surplus, therefore, gives the difference between what consumers *are willing* to pay (the total value they jointly attribute to *x* units of the product) and what they actually do pay.

Likewise, in section 2.3 we saw that the area beneath the supply curve from the origin to a specific quantity *x* equals the total variable production costs for that quantity. For this quantity producers receive an amount that equals quantity *x* multiplied by the price per unit. Therefore the total sum received by the producers jointly equals the area of the rectangle opSx beneath the price line. The difference between the total variable costs and the amount received by the producers, indicated by PS in Figure 4, is a measure of the advantages enjoyed by the producers, and is called the *producer surplus*.

Some economic manuals prefer to define the producer surplus as the difference between the sum received by the producers and their *total* costs. If we want to use this definition we need to deduct a constant amount, equal to the total fixed costs, from area PS. The difference in both definitions of the producer surplus does not lead to problems because we are usually interested in the *change* in the producer surplus as a consequence of an increase or decrease in production size, whereby the fixed costs remain unchanged.

The sum of the consumer surplus and the producer surplus is the *total surplus*. *It is a measure of the total economic welfare (or prosperity) of the market in question.*

At the point of intersection of the demand- and supply curves, when the marginal benefits and the marginal costs are equal, total surplus, and therefore welfare, is maximal. If the amount of produced and consumed goods is larger or smaller than the equilibrium amount, this will lead to welfare losses, as will be explained below.

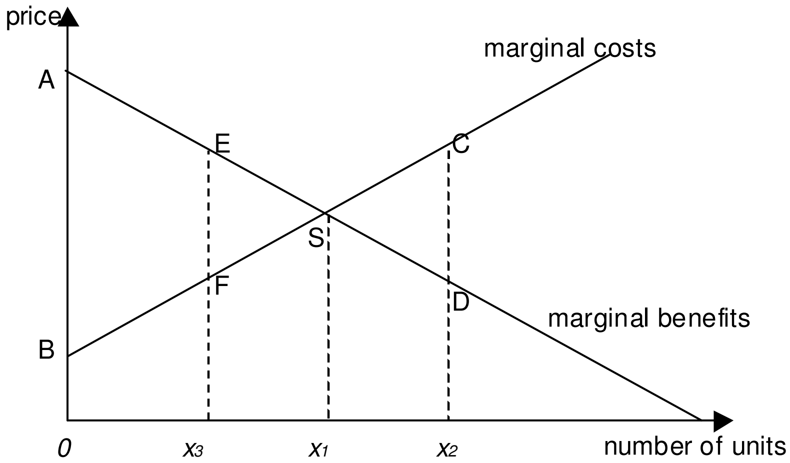

Figure 5 Efficiency loss due to overproduction or underproduction

The equilibrium amount in Figure 5 equals x_1. If this quantity is traded against the corresponding equilibrium price, total welfare, as explained above, equals the area of the triangle ABS. Now assume that the total amount of produced and consumed foods exceeds the equilibrium amount. This is called *overproduction*. Assume that the quantity traded is not x_1 but x_2. Marginal costs exceed marginal benefits between x_1 and x_2. There is a loss of welfare that equals the size of the area of the triangle SCD. Total surplus now equals the area of the triangle ABS minus the area of the triangle SCD. The area of the loss triangle SCD is also termed *dead-weight loss caused by overproduction*.

An analogous line of reasoning applies to *underproduction*, i.e. when the quantity traded is smaller than the equilibrium quantity. An increase in the traded quantity would, in that case, lead to an increase in total surplus because there still are consumers in the market who are prepared to pay above the marginal costs. In Figure 5 the area of the triangle SEF gives the *dead-weight loss caused by underproduction*.

In conclusion we can say that maximal welfare is realised at a production level where marginal costs equal marginal benefits. This is, in fact, also the equilibrium point that occurs in a perfectly competitive market. *Thus, a perfectly competitive market leads to the most efficient allocation of economic means, because in such a market marginal costs equal marginal benefits.*

2.7 Welfare loss caused by monopoly

A monopoly exists when a single producer controls the entire market for a particular product. A pure monopoly situation rarely occurs, nor does a perfectly competitive market. At most, we can say that a particular market tends to be competitive or, on the contrary, that is shows monopolistic tendencies.

Monopolies can arise in a number of ways. Legal restrictions, for example, can hamper market access. In other cases unique, carefully guarded technical knowledge, protected by additional patents, can prevent market access by other companies. Another example is the so-called natural monopoly, discussed later in this section.

Producers cannot influence price in a perfectly competitive market. They are *price takers* and need to take the market price as a given fact. The monopolist can be a *price setter* because he is the only supplier. In setting his price he starts from the demand curve. He can either choose a point on the demand curve characterised by a high price and a low volume, or he may decide on a low price and high sales. Maximal profits will determine his decision. Profit W equals total revenue (or turnover) TO less total costs TK:

$$W(q) = TO(q) - TK(q)$$

Here total revenue equals q multiplied by the willingness to pay for a quantity q that can be read from the demand function.

Profits will be maximal when the derivative of $W(q)$ equals zero, or when:

$$\frac{d(TO(q))}{dq} = \frac{d(TK(q))}{dq}.$$

The term on the left is called *marginal revenue*, the term on the right we recognise as the marginal costs. Maximum profits for the monopolist occur, therefore, when the marginal revenue equals the marginal costs.

For a linear demand function, the marginal revenue function is also linear, as can be easily verified[1]. For a non-linear demand function marginal revenue is given by a curved line. But whether the line is straight or curved, the marginal revenue function will always lie below the demand function. This can be derived analytically, but it can also be intuitively grasped. Though the sale of an extra unit yields extra revenue equalling the price level at the current quantity, the producer must lower his price to sell that extra unit, which lowers the price of the previous units also.

[1] For a linear demand function total revenue equals q times a linear function in q. So total revenue is quadratic in q, which means that marginal revenue (the derivative of total revenue) is linear in q.

Now look at Figure 6. At the point of intersection F, marginal revenue equals the marginal costs. In order to maximise his profits the monopolist will restrict his output to a quantity *x* and sell at price *p*.

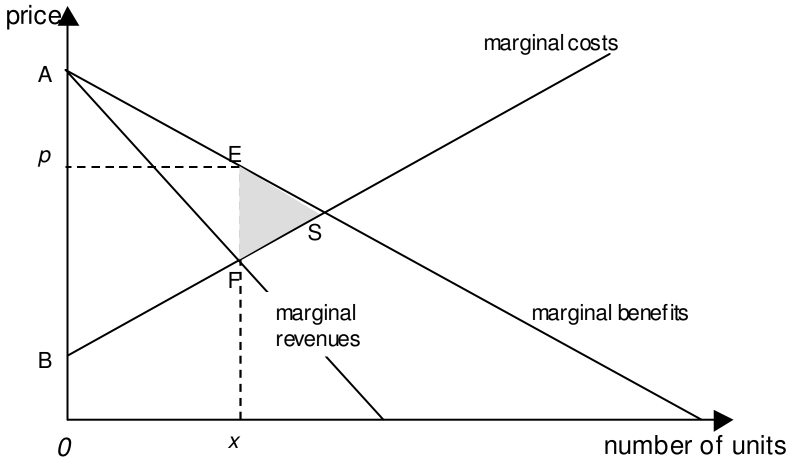

Figure 6 Pricing strategy of a monopoly

In a perfectly competitive market the total surplus would equal the area of the triangle ASB. In the case of a monopoly the area is reduced to AEFB. The result is a *dead-weight* welfare loss caused by underproduction equalling the area of triangle ESF. To enhance its own profit, a monopoly maintains production at a level that is too low, leading to inefficiency on this market. The situation is inefficient because, beyond the quantity *x*, there are consumers whose willingness to pay exceeds marginal costs.

In the transport sector, so-called *natural monopolies* are important phenomena. A natural monopoly comes into being when a single firm is able to offer a product or service to an entire market at a smaller cost than two or more firms could. This occurs when a company boasts continuously decreasing average costs over the relevant range of output. This happens particularly in capital intensive sectors. Excessive starting-up costs then inhibit other companies from entering the market. Public utilities such as public transport companies often find themselves in this situation.

A natural monopoly leads to additional complications caused by the decreasing average costs. The situation is shown in Figure 7. At decreasing average costs, the marginal costs always lie beneath the average costs, as explained under Figure 2 on page 11. In Figure 7, the MK line runs below the GK line.

The profit maximising monopolist would most like to confine his output to x_1, where marginal revenue equals marginal costs. This is a most undesirable situation, particularly for public utilities, because it leads to high prices, small outputs and welfare loss, as explained earlier.

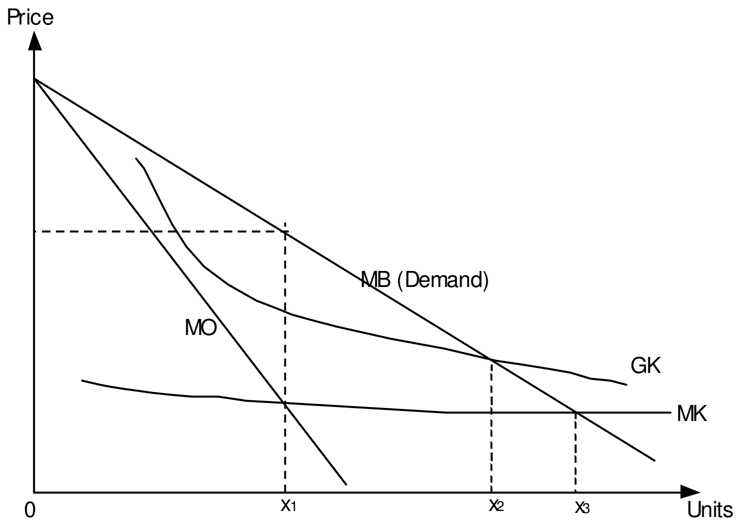

Figure 7 Natural monopoly

The ideal would be to produce output x_3, where marginal benefits equal marginal costs. The total surplus is, after all, at a maximum level when the marginal benefits equal the marginal costs, as explained in section 2.6. Governments could oblige public transport companies to realise an output of x_3 against a price that equals the marginal costs. The problem here is that the price would then be below the average costs. This means that total revenue would be below total costs, in which case the company would be loss making. Under these circumstances, a commercial company would not be able to survive.

If the company in question is a state company, subsidies could cover the company's losses. This argument is often used in defence of a subsidy-policy for loss-making public utility companies. Note: the company incurs losses because from a commercial point of view only the producer surplus counts, which is negative in this case. Extending subsidies, however, is advantageous for society as a whole. After all, seen from society as a whole the sum of the producer surplus and the consumer surplus matters.

Another solution would be to grant a company permission to apply *price discrimination.* The concept of price discrimination is explained in the next section.

A compromise-solution, lastly, is to oblige companies to realise an output of x_2 at a price that equals average costs. The advantage here is that a company breaks even and does, therefore, not need subsidies. Welfare losses do, however, still occur but are accepted.

A price-discriminating monopolist

Price discrimination, much used in transport, means that different groups of consumers pay different prices for the same product. This is how a monopolist can endeavour to increase his profit. The various transport tariffs that a railway company charges to its various customers is an example of this situation.

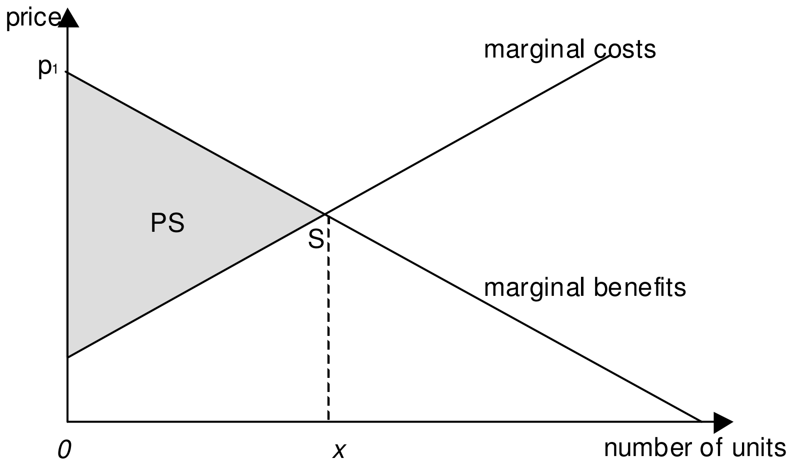

Figure 8 Perfect price discrimination

Perfect price discrimination means charging the maximum price that a customer is willing to pay. In Figure 8, the first customer's willingness to pay equals p_1. That is also the price he is charged. And so on. Every customer who is willing to pay more for the product than the marginal costs is charged a price equal to what he is willing to pay, as shown by the demand curve. As we know, the total surplus equals the shaded triangle. Note that the entire total surplus now consists of producer surplus, there is no consumer surplus. Also note that in the case of perfect price discrimination there are no welfare losses.

Since most people do not make their willingness to pay apparent, we must find a way in which to distinguish consumers who differ in their willingness to pay. An absolute distinction will be impossible but a number of different ways are used to enable a degree of group-distinctive patterns to emerge. Tariffs can be fixed according to the age or the status of customers. Young people pay less for their train fare than older people do and businessmen pay more than students do. Quantum discounts are another method. In the aviation sector, where price discrimination is widespread, lower prices are charged for journeys that include a weekend away. A businessman with a higher willingness to pay will not be too interested in this offer, while it may be very acceptable to a tourist with a lower willingness to pay.

2.8 Welfare loss caused by externalities

Externalities are negative or positive side effects that incur costs or provide benefits to third parties or bystanders, i.e. people other than those directly involved in the production or consumption of a certain good. Parties involved in the economic transaction do not take these side effects into consideration, which means that the side effects are not reflected in the prices. External benefits are positive side effects; external costs are negative side effects.

There is a distinction between *private* and *social* benefits and costs. Private means that we look exclusively at the benefits and costs of the parties directly involved in the economic transaction. If we include outsiders as well, in principle society as a whole, we speak of social benefits and costs.

A *consumption externality* means that there is a difference between private and social marginal *benefits*, in other words if there are external effects in consumption then there is a difference between the private and the social demand curve.

A *production externality*, on the other hand, deals with an underestimation, respectively an overestimation of the social marginal *costs*. From now on we confine ourselves to the negative externalities that occur in production, and in particular the external costs of private transport. Here we have an *underestimation* of the social marginal costs.

One always hears of the external *costs* inherent in the utilisation of infrastructure. But are there any external *benefits* connected to such utilisation? Some people say that they do exist and they refer, for example, to the positive effect of transport on the economic development of a region, increased employment, lower prices for goods and services and increased leisure-time. But these effects are not *directly* linked to the use of infrastructure. They are *indirect* effects linked to the construction of infrastructure. We return to this subject in section 4.5 where we shall see that the above benefits are rarely additional to the direct benefits, but represent, in fact, a redistribution of the original direct infrastructure benefits (consisting of timesavings). In other words: the indirect effects are already implicitly present in the demand curve.

To qualify as an external benefit, the *direct* positive effect of his trip should not be taken into consideration by a road-user when deciding whether or not to undertake his journey. Or, by illustration, of what advantage is it to you or anyone else if your next-door neighbour decides to take the car to work? There is common agreement amongst economists that the use of infrastructure does not deliver direct external benefits of any significance.

There are, on the other hand, a large number of external costs involved in the use of infrastructure. More specifically these are the costs incurred by loss of time caused to others, the potential costs of increased risk of accidents that involve others and the costs of noise pollution and environmental pollution that hinders others.

When assessing the welfare effects we need to start from the costs imposed on society as a whole: the *marginal social costs* (MSK). A consumer bases his decision on his own personal costs alone: the *marginal private costs* (MPK). *Marginal external costs* (MEK) represent the difference between the marginal social and the marginal private costs:

$$MSK(q) = MPK(q) + MEK(q)$$

The price on the market for a product that causes negative production externalities is actually too low from the social point of view. This is because the costs to society as a whole have not been included in the price decision of the producer. Therefore the supply curve or the marginal cost curve is too low and the intersection with the demand curve leads to a level of production that is too high. This *overproduction* leads to social welfare losses. See Figure 9.

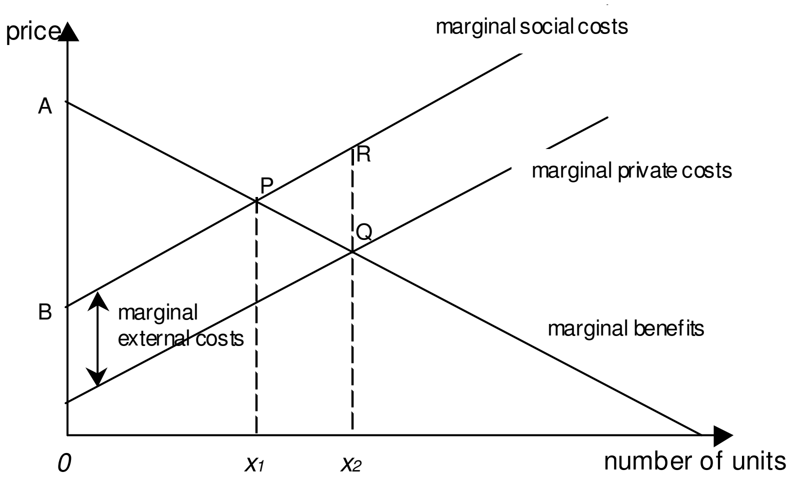

Figure 9 Welfare loss caused by neglecting external costs

The efficient quantity that should have been produced is x_1, at which the marginal social costs equal the marginal benefits. In this situation the total surplus equals the area ABP. In reality the producer only takes into account his marginal private costs and produces x_2 units. Beyond x_1, however, the total marginal costs for society exceed the marginal benefits, which represents a loss equal to the area PQR. In this last case, therefore, total surplus equals ABP minus PQR.

If externalities lead to a failure of the market to achieve an efficient market equilibrium, the government can intervene, for example by introducing taxes (tolls) proportional to the external costs. In such cases transport consumers will take the extra cost caused by their actions into consideration. This is called *internalising the external costs*.

Positive production externalities can also cause a loss in welfare! In such cases the market price is too large when seen from the social viewpoint. The private supply curve is too high and the point of intersection with the demand curve leads to a quantity that is too *small*. Such *underproduction* leads to social welfare losses

3. Utilising the infrastructure: optimal pricing

We begin this chapter with an analysis of the current market equilibrium in road traffic. The analysis will show that the current market equilibrium is inefficient and that it leads to welfare losses. Pricing mechanisms should lead to more favourable market equilibrium. Next we will examine the distribution of the welfare revenues gained under an optimal equilibrium regime. In order to get some impression of the extent of the external costs involved in the road traffic sector we will have a brief look at some of the results of a project that was carried out in Flanders by the faculty of Applied Economic Sciences at the KU Leuven. The introduction of general pricing on the complete road network might be difficult to implement. If this proves to be the case, one has to resort to so-called next best or second best solutions. What that means will be discussed in the second last section of this chapter. We conclude the chapter with a short discussion of a number of additional important aspects.

The text in this chapter is partly based on chapter 1 of De Borger en Proost[3]. We refer the reader to that work for more detailed information and for additional references.

3.1 Market equilibrium of road traffic

Consider a section of a motorway during a particular period, for example the road from Leuven to Brussels during the morning peak. Assume that x cars use the road during the peak hour. What are the *total* costs of the utilisation of that road by these x cars?

We consider all costs, not only the private costs borne directly by the driver, but also the external costs that are carried by third parties. These external costs are caused by the driver, but are not or barely considered by the driver when making his mobility decision.

Infrastructure construction costs

Before transportation can happen we need to invest in new infrastructure. The construction of infrastructure involves not only the immediate building costs but also the costs of infringements or harmful effects on the surrounding area. These infringements include visual hinder and loss of identity of (historic) landscapes, which lowers the perception value of that environment. There also is landscape fragmentation. The barriers caused by infrastructure projects negatively affect the quality of life and they are also seen as one of the most important causes of loss of bio-diversity in plants and animals.

We summarise these various costs and call them the construction costs A. It is important to note that these costs are not directly connected to road use. Once the road is constructed, the volume of cars does not alter the size of these costs. The construction costs A represent fixed costs that are not dependent on x. This also applies to management costs, such as soft shoulder maintenance, lighting and surveillance.

Cost of car ownership and car use

With car costs we mean the car acquisition costs and the costs of the use of the car. The car has to be manufactured. This cost is converted into the acquisition costs. Translated into the depreciation costs this gives us an amount per trip. Then there are also maintenance- and fuel costs. Assume that the total car costs (taxes excluded) are a Euro per car. For x cars these costs are $a.x$ Euro.

Taxation

Governments raise taxes on the acquisition of cars and on car maintenance. But car-use is also taxed through fuel levies and traffic taxes. We assume that the total tax payments are *b* Euro per car. This amounts to *b.x* Euro for the *x* number of cars.

Time costs

The trip takes a certain length of time. The time spent on the road (whether or not in traffic-jams) is lost time that could have been used more productive. This unproductive time of the joint *x* drivers, should, therefore, be entered on the debit side. This is done by the introduction of a time valuation factor expressed in Euro per time unit. The time valuation factor is often indicated by *VOT*, which means *value of time.* The value of time factor is often taken to be constant. But this need not be the case. It could, for example, depend on the duration of the trip itself.

The value attributed by people to an hour of travel time can vary depending on the trip purpose and on the person making the trip. It is plausible that there will be a connection between the individual's income and the value he places on his time. The value placed on time can be deduced from the behaviour of travellers in situations in which they can choose between long but cheap routes and shorter but more expensive routes. Home-work traffic shows values of around 7 Euro/hour. The values for commercial traffic can be up to 20 Euro/hour.

Increased flow rates on road sections lead to increased travel time. The link between traffic load *x* and travel time *t* is expressed by the travel-time function *t(x)*. Many such travel time functions have been developed. A much-used one is the so-called BPR-function, where BPR stands for Bureau of Public Roads.

When we multiply the value of the travel time function by the VOT-factor we get the time costs function *c(x)*. This function expresses that the travel time costs for one car are *c(x)* when there are *x* cars on the road. Time costs are *c(x)* for one car, and thus *x.c(x)* Euro for all cars together.

Environmental costs and other social costs

We already mentioned that the *construction* of infrastructure can lead to damage to the landscape and may negatively affect the quality of life. The *use* of roads also leads to environmental damage, namely atmospheric and other pollution. And other detrimental effects of road traffic, such as noise pollution and lack of safety should not be neglected either. Wear and tear of the road, lastly, incurs maintenance costs that will rise if the road is intensively used. Simplifying matters greatly, we assume that the environmental- and other costs are *m* Euro per car, and therefore *m.x* Euro for an *x* number of cars.

Total social costs and marginal social costs

The total social costs for all *x* vehicles together are:

$$TSK(x) = A + a.x + b.x + x.c(x) + m.x$$

The marginal social costs are found by differentiation of *TSK(x)* with respect to *x*:

$$MSK(x) = a + b + c(x) + x.\frac{d(c(x))}{dx} + m$$

Note that the infrastructure construction costs *A* are not part of this marginal cost function. These are fixed costs that do not increase with increased road use. The costs incurred by landscape fragmentation should not be included either since they were incurred at the construction stage and also represent a fixed cost. This is why these costs do not appear in the marginal cost function either.

This by no means implies that these costs are unimportant. They play a significant role in the decision making process for the construction of new infrastructure. They are important elements in a so-called costs-benefits analysis, a subject to which we will return in chapter 4.

3.2 The current market equilibrium

The marginal social costs are outlined in the diagram in Figure 10.

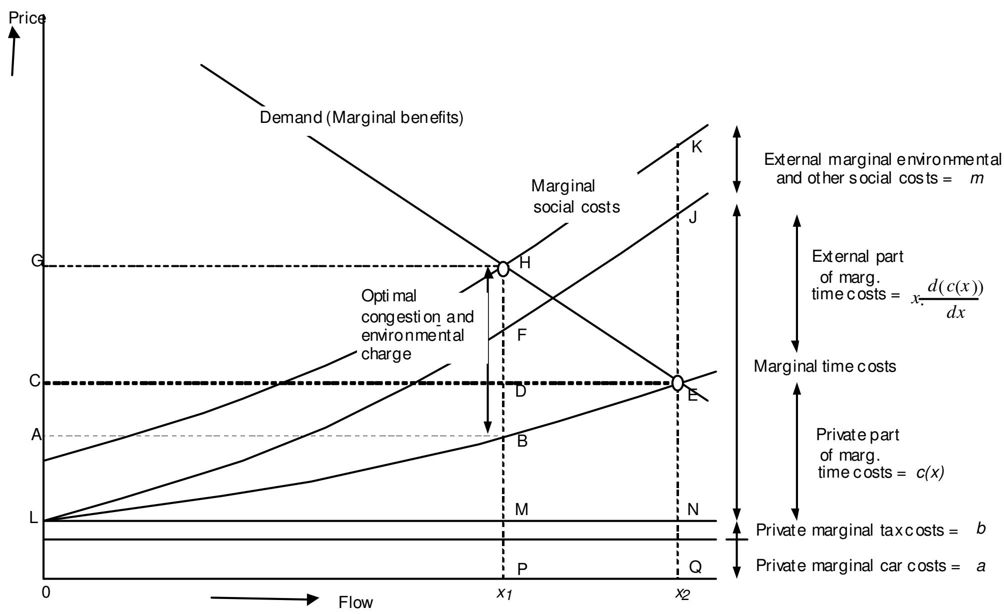

Figure 10 Existing and optimal market equilibrium for road traffic

They are equal to the sum of the private marginal costs, the marginal time costs and the marginal environmental costs. The demand function or the marginal benefits function represents the demand for vehicular traffic on the road section. The demand function gives the number of vehicles that want to use the road as a function of the generalised trip price. Few drivers are willing to make the trip when the price is high. As the price gets lower, more people are willing to use the road.

How many cars will now use the road? We usually find this quantity by determining the point of intersection of the demand function and the marginal cost function. However, when the driver determines his costs he does not consider all the terms shown in the social cost function *MSK(x)*!

The driver takes the following costs, his private marginal costs *MPK,* into consideration: the marginal car costs *a*, the marginal tax costs *b* and the travel time costs *c(x)* that are experienced by him.

$MPK(x) = a + b + c(x)$

The *MPK* curve cuts the demand curve at E. The volume of vehicles belonging to this equilibrium is x_2. Thus a number of x_2 cars will use the road.

3.3 The optimal market equilibrium

The driver, in his conduct, does not take account of the marginal external costs *MEK*. These are the marginal external environmental costs *m* and the external part of the marginal time costs expressed by the term *x.d(c(x))/dx*. Though he causes these costs, they are not charged to him, so that they do not affect his behaviour.

$$MEK(x) = m + x.\frac{d(c(x))}{dx}$$

The marginal external environmental cost *m* probably does not need further clarification. What, however, is meant by the external part of the marginal time costs represented by the term *x.d(c(x))/dx*?

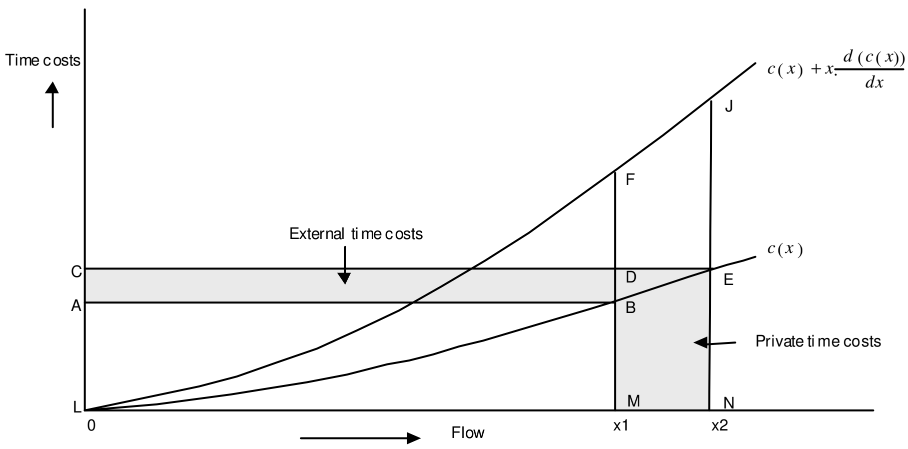

Figure 11 External and private time costs

This is explained in Figure 11. When a driver joins a flow of traffic consisting of *x* vehicles he is confronted with time costs equal to *c(x)*. He only takes these time costs, the private time costs, into account. He does *not* take account of the fact that the speed of the entire flow decreases slightly because he joins it. By his decision to join the traffic flow, travel time increases by an amount of *d(c(x))/dx*. This deceleration is experienced by all *x* cars in the flow, which explains the term *x.d(c(x))/dx*.

In Figure 11 the total time costs at a flow rate of x_1 equal the area ABML. At a flow rate of x_2 they equal the area CENL. Therefore, if traffic flow increases from x_1 to x_2 total time costs increase with an amount equal to the shaded area ABMNEC.

The increase in total time costs also equals the area beneath the marginal time costs function from x_1 to x_2, i.e. equal to the area FMNJ. From ABMNEC = FMNJ we deduce ABDC = DEJF. We will need this result later on in Table 2 on page 31, when we make a profit and loss account for the parties involved in road pricing.

The equilibrium E in Figure 10 is not desirable from a social point of view because the marginal social costs *exceed* the marginal benefits. The *optimal* equilibrium is the point where the marginal social costs equal the marginal benefits. This is point H in Figure 10. In this optimal market equilibrium, the number of cars using the road has decreased to x_1.

How can one adjust the existing market equilibrium E in order to arrive at the optimal equilibrium H? Drivers need to be made accountable for all the costs caused by them, including the external environmental costs and the external time costs. This can be done by charging a toll equal to the marginal external costs *at the optimal equilibrium*. This toll corresponds to the distance HB in Figure 10. Thus, the charge should not be equal to the external costs in the existing equilibrium x_2 ! If that were to be done, the toll would be far too high.

The transition from the current market equilibrium E to the optimal market equilibrium H produces a significant increase in welfare for society as a whole. As explained earlier in this chapter the increase in welfare corresponds to the area of the triangle HEK in Figure 10. One might think that a toll higher than HB would benefit society even more, because congestion and environmental pollution would decrease to still lower levels. This, however, is incorrect. It would lead to "underproduction" of mobility; the welfare gain achieved for society would not be maximal. From a social point of view, therefore, accepting a certain volume of congestion and environmental pollution is optimal.

3.4 Equity aspects of road pricing

Internalising external traffic costs through the introduction of tolls ('road pricing') has obvious advantages to society as a whole. This has been known for a long time. It was first put forward around 1925 by the well-known economist Pigou and since then many other economists have proposed it.

But the introduction of road pricing has met with much opposition from many sections of society. Introducing a new tax has never been popular. People tend to forget that toll revenues collected by the government become available again to society and that they could be used, amongst other things, to compensate any injured parties. When such compensation has been extended, a net welfare profit remains (equal to the area HEK in Figure 10) for the common good.

Road pricing will affect different participants in different ways. We shall make a profit and loss account for the different groups involved. There are four groups. Amongst the drivers we distinguish the group that pays the toll and continues to drive (in Figure 10 indicated by the range $0 - x_1$), and the group that decides to leave the car at home because they find the charge prohibitive (indicated by the range $x_1 - x_2$). We also distinguish the group of people suffering the effects of harmful external environmental- and other effects, and lastly there is the government that collects the toll.

The group of drivers that continues to drive primarily gains some time due to reduced congestion. For them, travel time decreases from NE to MB. The time gain for the entire group expressed in monetary units is represented by the rectangle ABDC in Figure 10. However, their travel costs have increased with HB, the taxation raised. Total toll charges are reflected by the rectangle ABHG. This results in a loss in welfare CDHG for this group.

The tolls collected go to the government, who sees its revenues rise with the amount ABHG.

Now we look at the group of drivers that decides to leave the car at home because the toll makes driving too expensive for them. They suffer a loss of benefits equal to PQEH. However, they save on private costs PQED. The total loss to this group, therefore, is given by the area DEH.

Finally, reduced traffic leads to reduced environmental costs and other harmful effects. The increased welfare for the people involved equals the area FJKH. Adding all gains and losses across the four groups results in a total increase in welfare corresponding to the area of HEK. We found this same result earlier. See Table 2.

Table 2 Gains and losses of al parties involved in road pricing

Group	**Profit/loss**
Continues driving (0 to x_1)	- CDHG
Stops driving due to toll (x_1 to x_2)	- DEH
Government	+ ABHG
Victims environmental degradation	+ FJKH
Total	ABHG - CDHG - DEH + FJKH = ABDC - DEH + FJKH = DEJF - DEH + FJKH Because: ABDC = DEJF (see Figure 11 and the explanatory text on page 30) = DEJF + FJKH - DEH = DEKH -DEH = HEK

The analysis above shows that drivers, both those who continue to drive after the imposition of the charge and those who stop driving, suffer a net loss in welfare as a consequence of the introduction of the congestion- and environmental charge. This explains the strong social opposition to the introduction of road pricing.

The car driver's losses are, however, more than compensated for by the revenues of the government and the reduced environmental damage. Increased welfare for society as a whole, to which drivers also belong, becomes apparent only when the government reinvests its revenues in society. We will now pay some attention to the redistribution problem of the toll revenues.

Let us begin by saying that the toll revenues need not necessarily be reinvested in the transport sector. It is the responsibility of the authorities collecting the charges to spend the money in the most useful way. It is possible that the money can be usefully returned to the transport sector, to deal with the external environmental effects or to be used for the construction and extension of the infrastructure, for example. But it is also possible that much larger returns can be achieved in other sectors of the economy.

In order to gain political support for road pricing the most obvious course of action would be to reinvest part or all of the collected revenues in the transport sector. In this way, part of the taxes raised could compensate for the environmental damage and other social costs still caused by the drivers who remain on the road.

Effective expenditure of the revenues, however, is not so straightforward and is the subject of wide public debate. In principle, the affected drivers could be fully compensated for their loss in welfare. The problem here is, however, that these drivers could 'reinvest' their tax refunds in mobility all over again.

Calculations show that the lion's share of the external costs is incurred by time losses caused by congestion. Though the drivers themselves experience the loss in time, the costs are actually borne by their employers who, one would expect, pass the costs on in the prices of their products. In this roundabout way society is still faced with congestion losses. It has been suggested that the revenues could be used to reduce general taxes, so as to benefit society at large. Others argue that the revenues should remain in the transport sector, to stimulate public transport, for example, or to improve infrastructure. It looks as if the public debate about a fair and equitable redistribution of congestion charges is by no means over.

3.5 Magnitude of the external costs in road traffic

Table 3 gives data pertaining to the real amounts involved in the external cost of road traffic. The data come from Proost[4], who calculated these data for the reference year 2005 based on an extrapolation from 1990. The figures apply to the car transport sector during peak hours in non-urban environments.

An expected growth rate in car traffic for 2005 and the introduction of 'cleaner' cars has been taken into account.

Table 3 External costs in road traffic (non-urban)

marg. costs	Unchanged policy (no charges)				Optimal congestion- and environmental charge	
peak car non-urban traffic 2005	marg. external congest. costs	remaining marg. external costs	total marg. external cost	private taxation costs	total optimal marg external costs (toll)	taxation incl. toll
Euro/person-km	0.72	0.05	0.77	0.07	0.26	0.33
index traffic intensity	100%				83%	

The figures in the table are remarkable. Note in the first place that the bulk of the marginal external costs consists of external congestion costs. We also see that when tax reforms are not introduced, the private taxation costs bear no relation to the external costs incurred.

The optimal marginal external costs (HB in Figure 10) amount to one third of the current marginal external costs. Traffic flows would, with the introduction of an optimal charge, decrease from 100 to 83%. But this is only achieved with a considerable increase in taxation. There would be a 33/7 or almost a five-fold increase in taxation levels as compared to the current rates! Though these high taxes would apply during peak hours only, their magnitude shows how politically sensitive they are.

3.6 Congestion charges and investment in additional infrastructure

An important part of the external costs is caused by congestion. If we were to increase road capacity these congestion costs would obviously decrease. However, expansion also has its costs: we need investment capital to increase capacity

In 1962 Mohring en Harwitz[5] proved that the revenues from an optimal congestion charge would be just sufficient to finance capacity extension, if a number of plausible conditions are met.

The conditions are:

- *Time loss is a function of flow rate divided by road capacity.* If x represents flow rate and y capacity, then the time costs are a function of the quotient of x and y and are represented by $c(x/y)$.

  This expression means that we assume that a doubling of traffic flow does not influence travel time provided road capacity, for example the number of lanes, is also doubled. This is approximately true for most roads.

- *The cost of capacity extension K is a linear function of the capacity:*
  $K = a.y$ with a a constant.

This means that a doubling of capacity, for example, will need a doubling in investment costs. This seems quite reasonable, but there are instances, for example in densely populated areas in Flanders, where costs increase disproportional with additional traffic lanes. This is called diseconomies of scale in capacity extension.

- *An optimal strategy of capacity extension is used.* By this we mean that investments in capacity extension must be such that the marginal costs of the capacity enlargement exactly equal the marginal savings in travel time for all road users:

$$\frac{dK}{dy} = \frac{\partial(x.c(x/y))}{\partial y}$$

We will now show that the revenues from an optimal congestion charge are just sufficient to fund the extension investments, provided that the three conditions mentioned above are met. The derivation below is a slightly adapted version of the derivation given in an article by Gomez-Ibanez[6].

The optimal congestion charge *H* per driver equals (see section 3.3):

$$H = x.\frac{\partial(c(x/y))}{\partial x}$$

For the time costs function we have $c(x, y) = c(t.x, t.y)$ for arbitrary t, i.e. if x and y are each multiplied by the same arbitrary number t, the value of the function does not change. Such functions of x and y are called homogeneous functions of degree 0. According to Euler's theorem for homogeneous functions, we then have:

$$x.\frac{\partial(c(x/y)}{\partial x} + y.\frac{\partial(c(x/y)}{\partial y} = 0$$

Inserting this expression in the formula for *H* we get:

$$H = -y.\frac{\partial(c(x/y))}{\partial y}$$

The optimal strategy for capacity extension implies that:

$$a = -x.\frac{\partial(c(x/y))}{\partial y} \qquad \text{or} \qquad \frac{\partial(c(x/y))}{\partial y} = -\frac{a}{x}$$

Inserting this expression in the formula for *H* then gives:

$$H = \frac{a.y}{x} \qquad \text{or} \qquad x.H = a.y = K$$

This states that the toll revenues from the x drivers just suffice to fund the capacity extension K.

This interesting result means that toll revenues can be used as an indicator that extension investments are required. If the toll revenues increase excessively, this is a sign that capacity extension should be considered. And, in that case, the toll revenues are just enough to fund the extension. In fact, the toll revenues are used to find the efficient scale of production (see page 10).

However, we repeat once again that government is in no way obliged to spend the toll revenues in the transport sector.

3.7 First-best en second-best

Market failures

We already stated that a perfectly competitive market leads to the most efficient allocation of economic resources in that market. This is because in that case the marginal benefits equal the marginal social costs.

Marginal benefits do not always equal marginal social costs. For example, we already saw that when it comes to efficiency, a *monopoly* represents an unfavourable market structure since prices in this form of market often exceed the marginal social costs. This results in a loss in welfare due to underproduction. We also saw that losses in welfare are suffered when *external costs* are not taken into account. In such cases marginal social costs exceed marginal benefits and we get overproduction. In some sense, monopolies and external costs disrupt the functioning of a competitive market. That is why we call them *market failures*.

There are other circumstances that can lead to market failure. Some goods or services, the so-called *public goods*, are not or hardly ever produced by the free market. Examples are public services such as police and fire brigade and goods such as street lighting, flood disaster works and also untolled roads and streets. It is characteristic for public goods that consumption cannot be confined to a particular group that pays for them. This is precisely why private firms are not prepared to produce these goods. A last form of market failure is caused by the *absence of information* about the real costs and benefits of a product.

When markets fail it is desirable that governments should intervene, with corrective measures. If external costs are involved, the authorities can impose charges, as shown above, while the behaviour of monopolies can also be influenced by taxation measures. If lack of information causes a market failure, authorities can issue information about the products or services via the media. Public goods, lastly, need to be paid for from tax revenues.

A world in which all markets are perfectly competitive and where all market failures have been corrected for is called a *"first-best"* world. The maximal economic welfare (or the maximal total surplus) that can be achieved in such a world is called the first-best optimum. A condition for a first-best optimum is that marginal benefits equal marginal costs on all markets. In principle the word "world" should be taken literally here since all markets influence one another to some extent. In practice, however, it is sufficient to look at a cluster of markets that deal in goods or services that are, to a certain extent, complementary or that can substitute one another.

Now assume that, for some or other reason, one of the markets does not meet the requirement of equality between the marginal social costs and marginal benefits. Should we then ensure that prices on all the other markets equal the marginal costs, thus achieving the best optimum possible? The somewhat surprising answer is: no. Lipsey and Lancaster formulated the principle of "second best" in a pioneering article published in 1956: "if at least one of the markets is unable to meet the optimum conditions, all other markets must also deviate from these optimum conditions in order to achieve the highest possible efficiency". They termed the maximum total surplus that can be achieved in such circumstances a "*second-best*" optimum.

The concepts first-best and second-best are also used if only one maket is involved. If on that market marginal benefits equal marginal costs, then on that market the first-best optimum is realised. If, because of restrictions in traffic policy, this maximum surplus cannot be achieved, we have to see what is the best achievable result under the given restrictions. This is also called a second-best optimum for that one market.

The theory of second-best is actually a translation into economical terms of a well-known fact from mathematical optimisation theory. Assume that the state of a (economical or other) system can be described by a number of variables. Assume, first, that there are no restrictions to the values that these variables can assume. The first-best, or most desirable (in any sense), state of the system is described by certain values of the variables. Now assume that some of these variables, for some reason, cannot assume the value necessary for the first-best situation. Keep these variables at a constant value and again determine the most desirable state of the system in terms of the remaining variables. Call this a second-best state of the system. In general we will then find that for this second-best state the non-restricted variables have a value different from the value they had in the first-best state.[7]

<u>A transportation example</u>

Let us apply the above to the markets of car trips. In the case of a car-network every origin-destination pair is a separate market with a certain demand function. In the following we first look at the case of only one market, i.e. a network with only one origin-destination pair. After that we will briefly discuss networks with more than one origin-destination pair.

One OD-pair (one market)

To illustrate how a second-best optimum can deviate from a first-best optimum we will look at the classical “two-route problem”. Here drivers travelling from a given origin O to a given destination D can choose between a toll route and a route where, for whatever reason, no toll is or can be raised and which is, therefore, toll-free.

In practice one will often meet with such situations. Raising tolls on all links of the network could, for instance, be too costly. Another reason is that some countries prefer toll-free alternatives to routes on which toll is raised. Lastly, the introduction of a general toll would require a long time to implement in which a dual system of toll roads and toll free roads would have to exist side by side.

The example is taken from Verhoef et al[8].

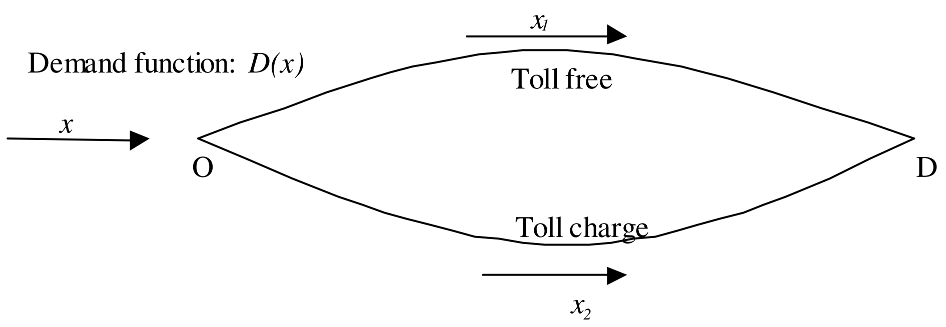

Figure 12 Toll charging with an alternative toll free route

We examine a single demand function *D(x)* for transport from O to D. Here *x* represents the total flow from O to D. This flow is divided over the two routes: $x = x_1 + x_2$. The demand function is:

$$D(x) = 50 - 0.01 \cdot x$$

where *D(x)* stands for the time costs for the trip from O to D. The number of cars *x* travelling from O to D increases as the trip costs *D(x)* decrease.

The private travel costs over both routes are given by the functions $c_1(x_1)$ and $c_2(x_2)$, both of which are also expressed in monetary units. In our example we assume that both cost functions are identical:

$$c_1(x_1) = 20 + 0.02 \cdot x_1 \quad \text{and} \quad c_2(x_2) = 20 + 0.02 \cdot x_2$$

For simplicity's sake, we only look at time and toll costs and ignore the vehicle and environmental costs.

According to the first principle of Wardrop, the user costs, i.e. private time costs and possible toll charges, along the two routes equal each other in the equilibrium situation. They also equal the value of the demand function *D(x)*.

Before determining the second-best solution, in which no toll can be raised on route *1*, we examine two other cases. In the first case, no toll is raised at all; the second case calculates the maximal welfare gain in the first-best situation when the optimal congestion charge applies on *both* routes.

The calculation of cases a) and b) is graphically illustrated in Figure 13

a) No toll on either route

Here we have:

$$c_1(x_1) = c_2(x_2) = D(x_1 + x_2)$$

$$20 + 0.02 \cdot x_1 = 20 + 0.02 \cdot x_2 = 50 - 0.01 \cdot (x_1 + x_2)$$

which leads to:

$x_1 = x_2 = \underline{750}$
$x = \underline{1500}$
$c_1(x_1) = c_2(x_2) = D(x) = \underline{35}$

Total surplus S equals the area beneath the demand function less total congestion costs across routes *1* and *2*:

$$S = \int_0^x (50 - 0.01 \cdot z)dz - x_1 \cdot c_1(x_1) - x_2 \cdot c_2(x_2) =$$

$$[50 \cdot z - 0.005 \cdot z^2]_0^{1500} - 750 \cdot 35 - 750 \cdot 35 = 63750 - 26250 - 26250 = \underline{11250}$$

b) "First-best" solution: congestion charges allowed on both routes

If toll is allowed on both routes then it is possible to reach a first-best situation, i.e. the situation where marginal benefits equal marginal costs.

We want to maximise total surplus. The total surplus equals the area underneath the demand function minus the total costs. The total costs are the congestion costs on route 1 and route 2. So the problem may be stated as:

$$\max_{x_1, x_2} S = \max_{x_1, x_2} \int_0^x D(z)dz - x_1.c_1(x_1) - x_2.c_2(x_2)$$

subject to the following conditions:
$D(x) = c_1(x_1) + t_1$
$D(x) = c_2(x_2) + t_2$

Because we are free in our choice of t_1 and t_2 there are no restrictions on the values of x_1 and x_2. By manipulating the “switches” t_1 and t_2 we can give x_1 and x_2 any value we want (within the range where both are positive and where $D(x_1 + x_2)$ is positive.

Maximisation means taking the derivatives of S (both with respect to x_1 and x_2) and equating them to zero. Because the derivative of the integral of the demand function equals the demand fnction itself and because the derivative of the total costs equals the marginal social costs, this amounts to the same as the rule stating that marginal benefits should equal marginal social costs for a first-best optimum. (When taking the derivative of the integral of the demand function, note that taking the derivative with respect to x_1 is the same as taking the derivative with respect to x because an infinitesimal change in x_1 gives the same change in x.)

Applying the rule above we find:

$$D(x) = c_1(x_1) + x_1.\frac{d(c(x_1))}{dx_1} \quad \text{en} \quad D(x) = c_2(x_2) + x_2.\frac{d(c(x_2))}{dx_2}$$

These equations state that the toll on both routes between O and D should equal the marginal external congestion costs on the routes in question. Note that these are the marginal external congestion costs in the equilibrium resulting after charging the tolls!

Applying this to our problem we find:
$50 - 0{,}01.(x_1 + x_2) = 20 + 0{,}02.x_1 + 0{,}02.x_1$
$50 - 0{,}01.(x_1 + x_2) = 20 + 0{,}02.x_2 + 0{,}02.x_2$

leading to:
$x_1 = x_2 = \underline{500}$
$x = \underline{1000}$
$c_1(x_1) = c_2(x_2) = \underline{30}$
$D(x) = \underline{40}$
$t_1 = t_2 = \underline{10}$

The total surplus now equals:

$$S = \int_0^x (50 - 0{,}01 \cdot z)dz - x_1 \cdot c_1(x_1) - x_2 \cdot c_2(x_2) =$$

$$[50 \cdot z - 0{,}005 \cdot z^2]_0^{1000} - 500 \cdot 30 - 500 \cdot 30 = 45000 - 15000 - 15000 = \underline{15000}$$

An alternative method to find the first-best solution runs as follows:

For the OD relation (market) as a whole the following time cost function applies: $c(x) = 0{,}01.x + 20$. We can find this time cost function by adding the cost functions $c_1(x_1)$ en $c_2(x_2)$ horizontally as indicated in **Error! Reference source not found.**.

The total costs for x car drivers thus equal $x.c(x) = 0{,}01.x^2 + 20.x$ and the marginal social costs (the derivative of the total costs) then come to:

$\text{MSK}(x) = 0{,}02.x + 20$

The marginal benefits (or the demand function which is the same) are:

$\text{MB}(x) = \text{D}(x) = 50 - 0{,}01.x$

Equating the marginal social costs to the marginal benefits then leads to the same values for x_1, x_2, t_1 en t_2 as calculated above.

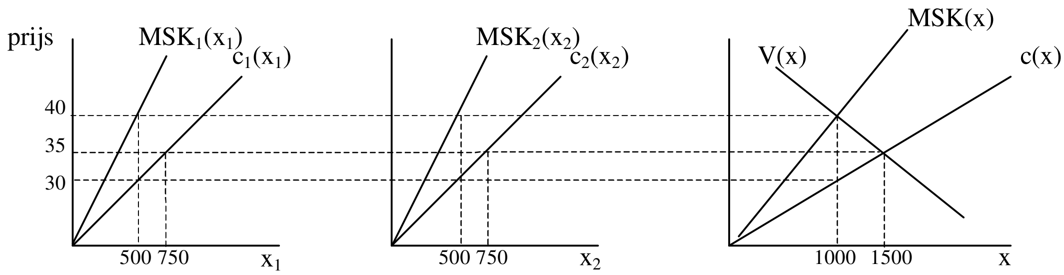

Figure 13 Calculation of the "two-route problem" for the no-toll situation and the first-best situation

c) Second-best solution: toll charging on route 2 only

User costs via route *1* are now $c_1(x_1)$ and via route *2* $c_2(x_2) + t_2$, where t_2 is the toll charged on route 2.

We want to determine the toll t_2 in such a way as to maximise the total surplus. The total surplus *S* equals the area beneath the demand function minus the total congestion costs. The problem therefore can be stated as:

$$\max_{t_2} S = \max_{t_2} \int_0^x D(z)dz - x_1 \cdot c_1(x_1) - x_2 . c_2(x_2)$$

under the following conditions:

$D(x) - c_1(x_1) = 0$
$D(x) - c_2(x_2) - t_2 = 0$

In contrast to the first-best problem, the fact that t_1 now must equal zero, imposes restrictions to the values that x_1 and x_2 can assume. The degree of freedom is one less. The boundary conditions are important now and limit the maximum value that S can attain

The method of Lagrange is the standard method to solve this type of optimisation problem. Since the demand- and cost functions are simple linear functions, we opt for an alternative method. We will express the objective function *S* exclusively in the variable x_1, whereupon we can establish the optimal value of x_1 by differentiation. Once the value of x_1 is known, x_2 and t_2 are easily found.

$c_1(x_1) = D(x_1+x_2)$
$$20 + 0.02 \cdot x_1 = 50 - 0.01 \cdot (x_1 + x_2)$$

And thus:
$$x_2 = 3000 - 3 \cdot x_1$$
$$c_2(x_2) = 20 + 0.02 \cdot x_2 = 80 - 0.06 \cdot x_1$$

After insertion of the expressions for x_2 and $c_2(x_2)$ we find:

$$S = \int_{0}^{3000-2\cdot x_1} (50 - 0.01 \cdot z)dz - x_1 \cdot (20 + 0.02 \cdot x_1) - (3000 - 3 \cdot x_1) \cdot (80 - 0.06 \cdot x_1)$$

After some computation we find:

$$S = -0.22 \cdot x_1^2 + 360 \cdot x_1 - 135000$$

The total surplus S is maximal when:

$$\frac{dS}{dx_1} = -0.44 \cdot x_1 + 360 = 0$$

therefore:

x_1 = 818.2 x_2 = 545.5 x = 1363.7
$c_1(x_1)$ = 36.4 $c_2(x_2)$ = 30.9 $D(x)$ = 36.4

The toll charge t_2 = 36.4 – 30.9 = 5.5 and S = 12273

The results of the calculation are summarised in **Error! Reference source not found.**:

Table 4 Results of the example calculation "two route problem"

	x_1	x_2	x	$c_1(x_1)$	$c_2(x_2)$	$D(x)$	t_1	t_2	S	ΔS
Without toll	750	750	1500	35	35	35	0	0	11250	
First best	500	500	1000	30	30	40	10	10	15000	3750 (100%)
Second best	818	546	1364	36.4	30.9	36.4	0	5.5	12273	1023 (27%)

The last column of the table shows that the first-best solution gives the largest welfare gain, compared to the situation without toll. In the second-best solution, where toll can only be charged on one of the two routes, the welfare gain amounts to a mere 27% of the gain in the first-best solution. The optimal toll in the second-best solution is only 5.5, which is significantly lower than the toll of 10 in the first-best situation.

In general, the second-best solution in the "two-route problem" is largely dependent on the time cost functions for the two routes and the elasticity of the demand curve. It is hard to draw general conclusions. In some cases, the welfare gains from the second-best solution can approach those of the first-best solution. In other cases (as in the example above) the second-best solution is significantly worse than the first-best solution. More information, also concerning the pricing of complex networks, can be found in Verhoef[9].

When travellers have a choice between transport by car or by public transport there is a problem that is somewhat similar to the "two-route" problem. In the first-best solution both car traffic and public transport should pay for their marginal costs. This may mean charging both cars and public transport! But say that, for whatever reason, for example because it is politically unfeasible, we cannot charge cars. In that case, the second-best solution may dictate that it is advantageous to extend a subsidy (a *negative* toll) to the public transport sector. That means that public transport is supplied at a price below the marginal costs. This entices travellers away from car transport, saving on external costs. A complicating factor in this situation is that new drivers could possibly take up the extra space on the road network.

More than one OD-pair (more interacting markets)

Realistic traffic networks generally consist of a very large number of OD-pairs. Every OD-pair in this case may be considered as a separate market with a corresponding demand function. As the paths between different OD-pairs also partially overlap these markets interact intensively.

In the discussion of the "two-route problem" it appeared that for a first-best solution on both routes the toll should equal the marginal external congestion costs on the route in question. For a large complicated network it can be shown that the rule also applies: the first-best solution consists of charging a toll on every used route between every OD-pair, where the toll on a route equals the sum of the marginal external congestion costs on all links that are used for the route. This, in turn, can be realised by charging on each link a toll equal to the marginal external congestion costs on that link (i.e. the marginal external congestion costs that apply to the equilibrium situation with tolls!). Because charging a toll on every link of a network is almost impossible, a first-best optimum will hardly ever be attained in practice. Calculating the first-best optimum however is useful to have an indication of the maximum attainable welfare gain.

Finding second-best solutions for realistically sized traffic networks is a topical research subject. One of the problems is finding the optimal toll given the links where tolls are allowed. Another, more difficult problem, consists of finding optimal locations for toll-installations subject to a limited available budget and possibly a restricted number of links where tolls are allowed.

3.8 Concluding remarks

In this section we summarise a number of important aspects linked to infrastructure pricing.

<u>Relationship between transportation markets</u>

In section 2.4 we indicated that a large number of market segments could be distinguished within the transport sector, segments that are defined in terms of transport relation, transport mode, trip purpose and timing of the trip. Timing refers particularly to the difference in peak- and off-peak hours.

In our discussion of pricing, we have basically confined ourselves to one single market segment, for example the market for car trips between Brussels and Leuven in the morning peak. However, in the transportation sector, no market segment can be looked at in isolation from other market segments. This is because the demand on one market depends on the prices of other markets that are trading similar products.

For example, if we introduce peak hour charges, part of the traffic will shift to the off-peak hours, which will increase demand in these off-peak periods. In addition, some car drivers will choose to travel by public transport. In short, there will be an interaction between connected market segments.

In section 2.5 we saw that these interactions can be described using cross-price elasticities. Computations may become very intricate and choosing the correct price-setting mechanism will require the use of mathematical models. An example of this kind of model is the TRENEN-model discussed in part 2 of De Borger en Proost[3], to which we refer for further information.

Increasing the variability of car costs

A proposed alternative to external cost charges is the introduction of more "variability" in existing car-related taxes. At present, car use is subject to a dual tariff: fixed charges such as a traffic tax and VAT (value added tax) on the acquisition of new cars and variable taxes pertaining to car-usage, primarily fuel-taxes. Car-use could be discouraged through the introduction of greater proportionality between taxes and car-use, i.e. through a shift from fixed costs to variable costs. A major disadvantage of such measures is that this kind of charge is not differentiated in time and space and is not linked, therefore, to the congestion caused by drivers. And it is exactly the congestion that causes most of the external costs.

Emission reduction through technological requirements to vehicles

In the discussion of external cost charges we added congestion charges and environmental charges together. It is inefficient, however, to reduce pollution caused by cars only by reducing the volume of traffic. There also is a need for technological minimum vehicle standards (the installation of converters for example).

4. Construction of new infrastructure: investment analysis

4.1 Social cost-benefit analysis

Both private firms and governments are occasionally faced with decisions regarding the desirability of investments into new means of production.

When a company contemplates the desirability of new investments, profitability comes first. A commercial enterprise usually begins by listing alternative possibilities for investment, where the zero-alternative (no new investment) is also an option. Then the firm will examine the consequences of each alternative in terms of costs and returns. The initial large investment expenditure will be compared to the benefits (turnover minus costs) realised over the life span of the capital good. The company will only decide in favour of new investment if the benefits exceed the costs, and only when it has become clear that investing the capital in an alternative way will not lead to greater returns.

Governments essentially follow the same procedure when evaluating the consequences of a public sector project. A project can, for example, involve the construction or the extension of transport- or other infrastructure or investment of funds into the health sector. A cost-benefit analysis for a private company is called a *private* cost-benefit analysis. When the government initiates the analysis on behalf of society we speak of a *social* cost-benefit analysis.

The most important difference between a private and a social cost-benefit analysis is the much broader view one takes when conducting a social cost-benefit analysis. A private investor only considers the expenditure and income of the company itself. Governments, on the other hand, consider the effects of the investment on all citizens. The most immediate beneficiaries of a new road are, naturally, the road users. But the interests of the people living in the surrounding areas, who may experience noise- or other pollution, must also be taken into account, as must the interests of the taxpayers who will eventually have to finance the construction- and maintenance costs.

The most characteristic feature of a social costs-benefits analysis is the fact that all effects, including the hard to quantify effects of environmental disturbance, noise pollution and so on, are expressed in one common measure. For practical reasons the monetary unit is chosen for this purpose.

In this section we give a brief overview of the main characteristics of a social cost-benefit analysis. Methods other than social cost-benefit analysis to assess government investments have been proposed. However, we do not dwell on these alternative methods, but refer to the literature, see for example De Brucker et al[10].

4.2 Steps in a social cost-benefit-analysis

The usual steps in a cost-benefit-analysis are the following:

- Establish the objectives of the project.
- Identify the project alternatives (including the zero alternative, also known as the base case).
- Set the analysis period, the geographical framework and decide on the discount rate.
- Analyse the direct and indirect project effects.
- Determine benefits and costs relative to the zero alternative.
- Evaluate possible risks.
- Present the results and make recommendations.

The analysis starts with a description of the *objectives* of the project. A clear description is necessary in order to limit the number of possible project alternatives.

The next step is to define the *project alternatives*. This process starts with the development of a so-called *zero alternative*. The zero alternative is sometimes called the "do nothing" alternative, but this should not be taken literally. The zero alternative means a realistic continuation of the existing situation without large investments but with continued normal maintenance works and small adjustments where and when required. The zero alternative is the point of reference to which the costs and benefits of the project alternatives are compared.

The project effects are compared to each other over a relatively long *analysis period*, also called the lifetime of the project. In principle the duration of the analysis period needs to continue as long as there are costs and benefits connected to the project. The normal analysis period for infrastructure projects is between 30 and 50 years.

The effects of improved infrastructure can be of local, regional, national or even international magnitude. Because of the so-called distribution effects inherent in the costs and benefits, it is important to define the geographical magnitude of the *area of analysis*. A project of national importance, for example, could have benefits that extend beyond the borders. The question is whether in that case the native taxpayers should be the only contributors to the investment.

After establishing the analysis period and the area to be analysed, a *discount rate* is decided on to compare present and future costs and benefits.

When analysing the project effects we need to distinguish between *direct effects* and *indirect effects*. The direct effects are the immediate transport effects experienced by the users and operators of the infrastructure. These will usually emerge as reduction in travel time for existing travellers and the extension of transport alternatives to new travellers. The impacts of transport on the environment and on liveability are also included in the direct effects. The direct effects, therefore, comprise the internal and external effects discussed in the previous chapter 2.8. The direct effects are usually estimated using traffic models. These estimations must, naturally, take the expected growth of traffic into account.

Infrastructure projects, however, do not only affect the immediate users and operators. Others are influenced also. Users pass their benefit on to third parties and therefore spread the original transportation benefit throughout the economy. These are the so-called indirect effects of an infrastructure project. *When determining the costs and benefits of a project, only the direct effects should, in principle, be included.* In most cases, the indirect benefits only represent the transfer of benefits and costs that have already been counted in the direct effects. If we were to add the costs and benefits of the indirect effects to those of the direct effects, this would amount to a *double count* of costs and benefits. Later in this chapter we will return to the important difference between direct and indirect effects.

The next step is to express the costs and benefits during the lifetime of the project in monetary units. The costs and benefits are determined relative to the zero alternative. Typical costs and benefits of an infrastructure project are:

- *Construction costs*. The construction costs include expenditure on design, expropriation, construction and maintenance. Although these costs are fairly simple to assess, they are often underestimated in practice.

- *User benefits and costs.* The most important benefit of an infrastructure project is usually the reduction in travel time as compared to the zero alternative. Other costs to travellers, such as vehicle- and fuel costs, are often also influenced by a project. The project may also lead to a reduction in the number of accidents adding to the benefits of the project.

- *External costs.* As explained above, externalities are the uncompensated direct impacts of the project on third parties. The increase or decrease in external costs relative to the zero alternative must also be taken into account. External costs include the costs caused by emissions, noise and other adverse effects on the quality of life. But they also include external congestion and accident costs. All these effects need to be expressed in monetary units.

There are many uncertainties in the evaluation of very long-term infrastructure projects. Future costs may be much higher than estimated, or future traffic volumes may deviate considerably from the volumes forecasted. This is why, especially in larger projects, a *risk analysis* is carried out. It tests the sensitivity of the results of a project for the values of the input parameters.

There are a number of ways in which to present the results of a cost-benefit analysis. The most frequently used indicators are the *net present value (NPV) and the benefit-cost ratio (BCR).* The net present value is the balance of all costs and benefits during the lifetime of the project, converted to their present value using the discount rate. The ratio of benefits and costs is an obvious criterion but is liable to misinterpretation. This is due to the fact that the ratio depends on the degree of aggregation of benefits and costs over the successive years of the project. Take, for example, a project that has costs 1000 units in years 1 and 2 and has benefits of 4000 units in year 2. If we simply add the benefits and costs we get a benefit-cost ratio of 2. However, if we aggregate the benefits and costs in year 2 to a benefit of 3000 units for year 2 we find a ratio of 3.

We will now take a closer look at some of the steps given above.

4.3 Discount rate and analysis period

Infrastructure projects usually have a very long lifetime. They tend to involve very high initial investment costs while the benefits manifest themselves in the long run. The costs and benefits must be valued depending on the moment they arise. A benefit of 1 Euro immediately available is worth more than a benefit of 1 Euro after five years. This is because the 1 Euro can be invested. At an interest rate of r % per annum the investment has grown to $(1+r)^i$ in i years.

This is why the various costs and benefits must be multiplied by a factor of $1/(1 + r)^i$, in which r represents the so-called discount rate per year and i the year in which the cost or benefit concerned becomes available. The balance of costs and benefits that emerges when the discounted costs are deducted from the discounted benefits for each year and then added over the lifetime of the project, is called the *net present value* of the project. The net present value is an amount expressed in monetary units that would, if it were to be paid out today, equal the total value of the project calculated over its entire lifetime. Execution of a project is justified if the benefits exceed the costs, or if the net present value is positive.

Table 5 gives an elementary example. Assume an initial investment in the first year of 30,000 Euro. The lifetime of the project is four years. Each of those four years yields a benefit of 10,000 Euro. Is this kind of investment justified? On the surface it appears to be so, for one spends 30,000 Euro and receives 40,000 Euro in return. However, this argument is too simplistic. The year in which the benefit becomes available is not taken into account, in other words a discount rate of 0% has been applied.

The table shows how the calculation should be done. A discount rate of 10% leads to a positive net present value. In that case, the investment is justified. A discount rate of 15 %, however, gives a negative net present value. Now the investment is not justified. The example shows that the choice of interest rate can affect the justification of an investment. A higher interest rate puts less emphasis on future benefits or costs. This is of importance for projects with possible harmful effects, for example on the environment, in the distant future. To avoid underestimating these long-term harmful effects, the discount rate should not be too high.

Table 5 Example of discounting

	not discounted		*discount rate 15 %*		*discount rate 10 %*	
year	*benefits*	*costs*	*benefits*	*costs*	*benefits*	*costs*
present		30,000		30,000		30,000
after 1 year	10,000		8,696		9,091	
after 2 years	10,000		7,561		8,265	
after 3 years	10,000		6,575		7,513	
after 4 years	10,000		5,718		6,830	
Total	*40,000*	*30,000*	*28,550*	*30,000*	*31,699*	*30,000*
net present value			*-1,450*		*+1,699*	

The example above only served to illustrate the principle of discounting and is, for several reasons, not representative for the kinds of investment usually involved in infrastructure projects. In real life, the amounts are usually much larger. The analysis period is mostly in the of 30 to 50 years range, while the discount rates used in the example are high compared to those used in practice.

In discussing the discount rate, we left the effects of inflation or deflation out of the calculation. Currency from which the inflation effect has been removed is called "real" (or "constant") currency. A real Euro buys the same amount of goods now and in the future. If we ignore the influence of inflation we are using a *real discount rate*. We can tackle the problem of inflation in two ways. We can apply a yearly inflation correction and apply the real discount rate, or we can integrate the expected inflation percentage in our interest rate. In the last case we use so-called nominal currency (a currency that maintains the same value in name only) and we apply a *nominal discount rate*. The usual interest percentages charged for long-term loans can be used for the discount rate. In many countries the discount rate applied to cost-benefit analyses is prescribed by the government. Table 6 below gives the real discount rates used in a number of countries.

Table 6 Real discount rate applied in a number of countries[10]

Belgium	4%	Spain	6%
The Netherlands	5%	Denmark	7%
Germany	3%	France	8%
The United Kingdom	8%		

An approximate formula is sometimes used to calculate the net present value. The net present value of a perpetual benefit and cost flow of a fixed yearly amount equals the fixed yearly amount divided by the discount rate. This can be proved quite easily using the formula for the sum of an infinite geometric progression. (Example: a yearly benefit of 1000 Euro against 8% yields a net present value of 1000/0,08 = 12500 Euro.)

Table 7 shows that the formula leads to an overestimation of the real net present value. The approximation improves as the discount rate increases and the analysis period is longer. Using the approximation formula is sometimes justified by the argument that there is often great uncertainty regarding the discount rate to be used and because of the arbitrary nature of the analysis period chosen. The residual value of a project is, moreover, often ignored, though it is potentially present.

Table 7 Exact and approximate calculation of net present value

Yearly benefit of 1000 Euro		*Net present value*		
discount rate	*analysis period*	*approximation*	*exact*	*overestimation*
8 %	50 years	12,500	12,233	2 %
8 %	30 years	12,500	11,258	11 %
4 %	50 years	25,000	21,482	16 %
4 %	30 years	25,000	17,292	45 %

4.4 Benefits and costs due to the direct effects

Figure 14 shows the distinctions made in the analysis of the effects of an infrastructure project. The direct (internal) effects are the transport effects (in particular the reduction in travel times) that are experienced by users and operators of the infrastructure. External effects (or externalities) experienced by third parties, such as environmental effects and the impact on the quality of life are also considered to be part of the direct effects. Indirect effects are the effects of the infrastructure project on the wider economy. In chapter 4.5 we will return to the indirect effects.

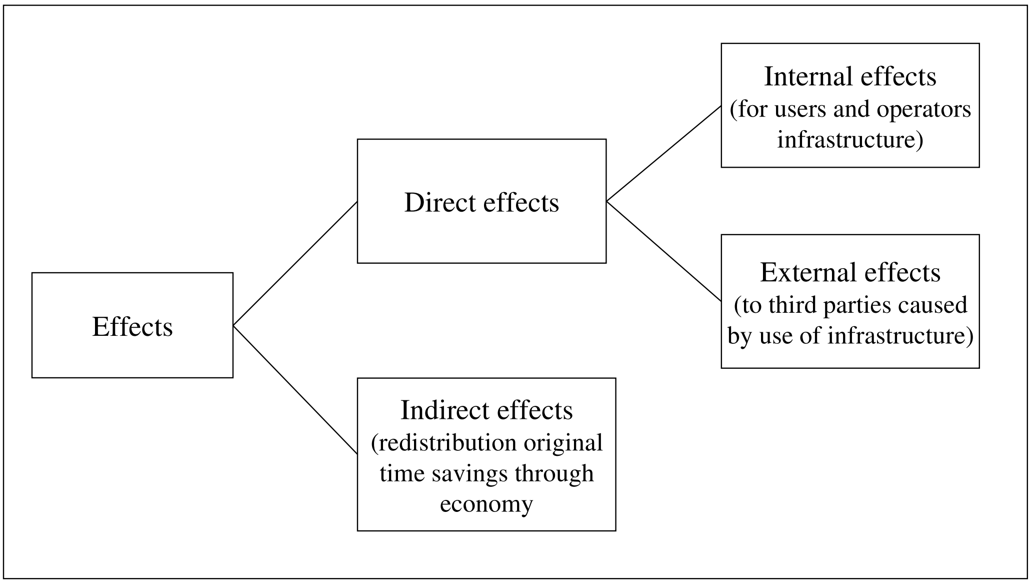

Figure 14 Direct and indirect effects of infrastructure

Internal effects

The direct internal effects will become particularly apparent as changes in traffic volumes and as changes in user transportation costs. Traffic volume forecasts can be made using traffic models. It must be kept in mind that changed traffic volumes do not only occur on newly constructed or extended infrastructure, but that new construction projects or extensions affect traffic volumes on the entire network. The cost-benefit analysis therefore needs to address itself to every part of the network that has been impacted by new construction.

The benefits of an infrastructure project are calculated using the demand- and supply functions. As we have explained elsewhere in this chapter, the area beneath the demand curve indicates the total benefits of a transport facility. The total variable costs are represented by the area below the supply curve. Therefore, if we want to find the costs and benefits relative to the zero alternative, we need only examine the differences in area between the zero alternative and the project alternative under consideration. We will elucidate this point by a simple example.

Say that there is an important road between two towns. The demand curve for that road connection in year j is indicated by Kj in Figure 15. The time costs function for the zero alternative (continuation of the existing situation without major investments) is c_0. In this example we assume that no toll is charged and that the private time costs alone determine driver behaviour. Traffic volume in year j, therefore, amounts to x_0 for the zero alternative with a corresponding generalised price of p_0 per trip.

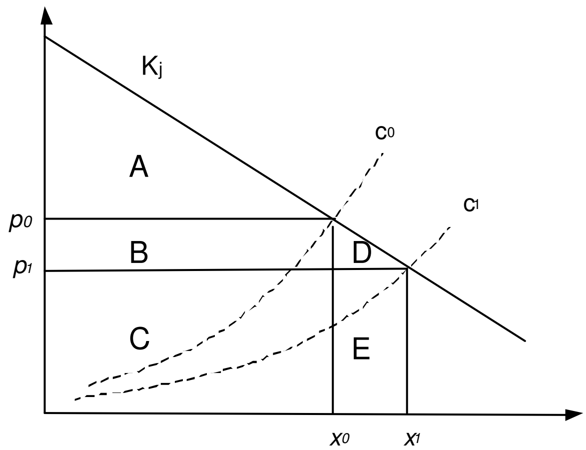

Figure 15 Calculation of the benefits of an infrastructure project

We will now examine the effects of a significant increase in the capacity of this road. The increase in capacity leads to lower generalised costs, given by the curve c_1. Thus, implementing this increase in capacity would lead to a volume x_1 and a generalised price of p_1 in year *j*.

The area A+B+C in Figure 15 indicates the total benefits for the zero alternative. Total costs equal B+C. (The generalised costs curves c_1 and c_2 represent average costs, not marginal costs! Therefore, it is not the area below the *c(x)*-curves that should be used to find the total costs. Instead, we should multiply the value of *c(x)* with *x*.) Thus, the total surplus for the zero alternative corresponds to area A. Equally the total benefits for the road with increased capacity equal the area A+B+C+D+E and the total costs are given by the area C+E. Thus, the total surplus in this case corresponds with A+B+D. Compared to the zero alternative, the capacity increase gives users an additional benefit of B+D in year *j*. Of this additional benefit, B goes to the existing road users and D represents the benefits to new users.

If we apply a linear approximation for the demand curve, we get the following expression for the area B+D:

$$0{,}5 \cdot (p_0 - p_1) \cdot (x_0 + x_1) .$$

This expression is sometimes termed the *rule of half.*

Remarks

- In the example above, we examined the usual calculation of direct benefits for an infrastructure project. In this example we assumed that external costs in the *use* of the infrastructure, be it in the existing situation or in a future situation, would not be charged to the road user. This has important implications for the results of the calculation.

  When users are not charged for the external costs, traffic volumes will be much larger than if they are charged. Because of the larger traffic volumes, more drivers gain time when new infrastructure is added. If congestion- and environmental charges are not applied, therefore, the benefits of an investment are higher than in the reverse case. Because of the high benefits there is a danger that an investment may be undertaken that would not have been justified if the infrastructure had been correctly priced. This is an additional argument in favour of the introduction of pricing mechanisms in transport. We refer to De Borger en Proost[3] for more details.

- The argument in Figure 15 was also based on the premise that demand would remain constant. However, for a proper calculation we need to know how demand will develop over the lifetime of the project. There are a number of aspects to be considered, for example changes in income patterns and changes in transport modes. But the most important factor is undoubtedly the growth in demand as population increases.

External costs

In the example above we only looked at savings in time costs. For a complete picture of the total costs and benefits accruing from an infrastructure project we must also look at the external costs relative to the zero alternative. These are the additional costs and benefits incurred by accidents, environmental damage and impacts on the quality of life. Note: when a project is assessed it is possible to include the external costs in a cost-benefit analysis, while at the same time accepting that these external costs are not passed on to the users through charges.

A characteristic feature of all cost-benefit analyses is that all effects are expressed in the same unit, i.e. in monetary units. For some costs such as construction costs this does not lead to any significant problems because available market prices can be used.

No market prices exist, however, for many of the costs inherent in larger projects, such as accident risks, environmental damage or noise pollution. Yet in a cost-benefit analysis these costs must also be expressed in monetary units. We will briefly examine a number of techniques developed by economists to valuate these project effects in monetary terms. In some cases, these valuation techniques are still highly disputed.

- *Accident costs*

  The benefit of a project may consist in a reduction in the number of serious accidents including fatalities. If we want to evaluate this benefit, we need to calculate the value of a human life. This is an extremely complicated and controversial problem. It involves not only medical costs and social production losses, but also the suffering caused by painful personal losses due to serious accidents. Valuation of human life is often based on the observed behaviour of individuals. It can be done by assessing what people are willing to pay to reduce the risk of an accident, for example by increasing the safety of their vehicle. Or one can examine the income differences between people involved in risky and less risky occupations.

- *Air pollution and climate change*

  Transport induced air pollution consists of harmful gasses that remain subsequent to the burning or evaporation of fossil fuels or solid particles that are directly emitted after burning. The emission of solid particles is a major problem in diesel engines. All emissions are particularly harmful to human health and they can harm the environment. Technical measures, such as the installation of catalytic converters, can significantly reduce most of these emissions. However, no technical solutions exist to reduce the emission of the CO_2 component. CO_2 emissions contribute to the greenhouse effect. Greenhouse gasses accumulate in the atmosphere and lead to increased mean temperatures on earth, possibly causing a rise in sea levels and climate change. An additional problem that has not yet been solved is the emission of extremely fine particles by diesel engines. These particles are very damaging to human health.

A number of techniques are used to value the cost of air pollution. These are based, for example, on the cost of purification techniques used by industry to reduce the concentration of emissions to 'acceptable' levels. Another method is to calculate the costs of medical treatment of complaints caused by the inhalation of polluted air. Yet another method is the so-called *hedonic price method*. This method aims to estimate the value of goods for which there is no direct market, by looking at the value of complementary goods. The behaviour on the housing market, for example, is observed. Price differences for otherwise comparable houses may be caused by environmental pollution. The magnitude of the price difference gives an indication of the value that people attach to a clean environment.

Valuation of the social costs of climate change is a perilous undertaking and extremely controversial. One method is to calculate the costs of the damage that will arise due to climate change or the costs involved in preventing that damage. Another approach involves assessing the costs associated with a reduction of CO_2 emissions to internationally accepted levels, as agreed, for example, in the Kyoto protocol.

Because of the difficulties involved in assessing them, the values proposed for the costs of air pollution and climate change show wide variation.

- *Noise pollution*

Noise pollution can be extremely invasive. Levels of 130 dB(A) and above can lead to intense pain and permanent ear damage. Levels between 40 and 90 dB(A) lead to sleep disturbance and levels below 40 dB(A) influence emotions and behaviour. Transport is the most important cause of noise pollution, more so than industrial noise or noise from neighbours. The monetary value of noise pollution is usually measured using hedonic prices.

- *Landscape quality*

The costs discussed above, namely those for accidents, air pollution and noise pollution are the most obvious costs. They play a role in both the assessment of the optimal use of infrastructure and in investment analysis using cost-benefit analysis.

A very important effect that should also be examined in an investment analysis (but not in pricing the *use* of infrastructure) is degradation of the quality of the landscape caused by infrastructure construction and the associated effects of visual interference and road barriers. Research into the valuation of these effects is at an initial stage.

Some data

Table 8 shows some data concerning the costs caused by externalities of road transport. They are taken from an extensive European research project[11] carried out in 1998. The costs involved are average costs. If costs are proportional to traffic flow, average and marginal costs coincide. This is certainly not the case for all categories of costs. Therefore, the values in Table 8 need to be seen as a rough indication of the marginal costs.

Table 8 Average external costs in road transport

	Private cars	Trucks
	Euro per 1000 vehicle kilometres	
Accidents	60	60
Noise pollution	5	23
Air pollution	13	66
Climate change	10	28

4.5 Benefits and costs due to the indirect effects

If a new road connection makes a district of a town more accessible, the firms in that district are liable to increase their prices. Rents of offices and homes will also increase. Travellers will initially profit because of reduced transport costs to the area, but in the long run their advantage will go to the firms in that area. Over time land prices will rise with the result that the aforementioned firms will have to pass their advantage on to the landowners. These and other mechanisms lead to a redistribution of the original infrastructure benefits.

Improved infrastructure may also give rise to a reorganisation of production. A faster connection gives haulage firms an initial cost advantage. However, sooner or later competition will force them to pass their cost advantage on to their customers. The laws of competition will dictate that these customers will, in their turn, have to relinquish their advantage through lower product prices. Thus, mutual transactions ensure that the infrastructure advantage is spread through the economic system at large.

The two examples above demonstrate the indirect effects of improved infrastructure. It can be proved[12,13] that in an economy where there is perfect competition in all markets the indirect effects merely lead to a redistribution of the material increase in welfare originally experienced by the direct users of the infrastructure. Redistribution does *not* generate additional welfare. In other words: the benefit of infrastructure improvement is in many cases (i.e. where the markets in question are sufficiently competitive) already implicitly included in the demand curve for transport.

Yet, although the indirect effects generally lead to a redistribution of welfare, there are cases where additional increases or decreases in welfare can arise. This happens when the infrastructure improvements affect markets that are subject to market failure. We already discussed the concept of market failure. In competitive markets marginal costs equal marginal benefits. Market failure occurs when the marginal costs are not equal to the marginal benefits. This leads to under- or overproduction of goods, which, in turn, causes *dead-weight* losses in welfare. Monopolistic behaviour is an important example of such a market failure. Markets also fail when external costs or benefits appear in an economic transaction.

Infrastructure improvement can sometimes reduce market failure in specific economic sectors, or it can lead to the opposite: an increase in market failure. In the first case the indirect effects lead to increased welfare, in the second case they lead to a loss in welfare, in both cases additional to the welfare gain caused by the direct effects. We will explain this by a number of examples.

A situation where indirect effects can yield additional benefits happens when improved infrastructure leads to a reduction in monopolistic behaviour. Lower transportation costs can lead to an enlargement of the trading area and thus a similar kind of company can enter into competition with a monopoly already operating in the area. The increased competition leads to a reduction in *dead-weight* loss and thus to additional gains to the economy.

Sometimes improved transportation options lead to the formation of agglomerations or clusters of activity. According to Small[6] cluster formation leads to additional benefits. Companies benefit by each other's proximity for a number of reasons: they may enjoy economies of scale, clusters promote the formation of a highly skilled labour force and firms can benefit from exchange of information. Cluster formation can, therefore, lead to additional gains, provided they represent new clusters of activity and not merely the transfer of activity from one place to another.

Infrastructure improvements can also cause indirect effects that increase costs and that need, therefore, to be deducted from the gains realised by the new infrastructure. This happens if the infrastructure project results in the attraction of industrial activity that causes high external costs because of harmful production processes.

4.6 Risk analysis

A risk-analysis is often carried out at the evaluation stage of a large infrastructure project. This type of analysis needs to highlight a) any possible eventualities, b) the probability at which such eventualities might occur, and c) the nature of the consequences should these eventualities occur.

A risk-analysis is usually carried out using a *sensitivity analysis*. The first thing to be established in a sensitivity analysis is which input variables are susceptible to uncertainty. This is in response to questions a) and b) in the paragraph above. The value of one of these input variables is then varied, and the impact of this variable on the end-result of the cost-benefit analysis established. All other input variables are kept constant. This is the avenue by which the answer to question c) is achieved.

If the sensitivity analysis shows that the value of a particular input variable is crucial to the results of the analysis, one can look for ways to reduce the probability of a change in the value of this input variable or one can try to find ways to minimise possible consequences.

There often are a number of input variables that can change simultaneously. In such cases sensitivity analyses can be carried out by inserting different values for the various input variables. However, the rapid increase in the number of cases to be considered is a great disadvantage of such a procedure, for it renders the interpretation of the results difficult. A probabilistic approach would be preferable.

A probabilistic approach involves a so-called Monte Carlo type of simulation. This type of simulation assigns an appropriate probability distribution to all input variables that are thought to have uncertain values. To calculate the output of the cost-benefit analysis a random value is chosen for all input variables based on their probability distribution. This procedure is repeated over and over again. The result is a probability distribution related to the output of the cost-benefit calculation, for example the net present value of the project. Assume that there are a number of alternatives that realise the project goal. Based on the probability distribution of the net present value of each of the alternatives, a well-considered decision can be taken regarding the execution of one of the alternatives.

List of figures

List of tables

References

[1] Ministerie van de Vlaamse Gemeenschap (2001), Departement Leefmilieu en Infrastructuur, Mobiliteitscel, *Ontwerp mobiliteitsplan Vlaanderen (Mobility Plan for Flanders)*

[2] Biz/ed (2003), Virtual Learning Arcade run by University of Bristol (www.bized.ac.be) *Tackling Traffic Congestion*, Information about the METS (Metropolitan Transport Simulation) model.

[3] De Borger, B. en Proost, S. ed. (2001), *Reforming Transport Pricing in the European Union, A modelling approach*, Edward Elgar, Cheltenham UK, Northhampton MA, USA

[4] Proost , S. (2000), *Mobiliteit economisch bekeken (Mobility from an economic viewpoint)*, Bijdrage aan Lessen voor de 21e eeuw (Lessons for the 21st century), Krachten voor de toekomst, Redactie Raymaekers, B. en A. van de Putte, Universitaire Pers leuven/ Davidsfonds/ Leuven.

[5] Mohring, H. en Harwitz, M. (1962), *Highway Benefits*, Northwestern University Press, Evanston

[6] Gomez-Ibanez, J., W. Tye en C. Winston ed. (1999), *Essays in Transportation Economics and Policy: a Handbook in honor of John. R. Meyer*, Chapter 4: *Pricing* by J. Gomez Ibanez, Chapter 5: *Project Evaluation* by K. Small, Brookings Institution, Washington D.C

[7] Lawrence B. Solum (2003) This intuitive description of the Theory of Second Best comes from Solum's *Legal Theory Blog*

[8] Verhoef, E.T. , Nijkamp, P. en Rietveld, P (1996) *Second-best congestion pricing: the case of an untolled alternative*, Journal of Urban Economics, 40(3) 279-302

[9] Verhoef, E. (2000) *Second-best congestion pricing in general networks*, Tinbergen Institute Discussion Paper TI 2000-084/3 (www.tinbergen.nl) , Amsterdam / Rotterdam

[10] De Brucker K., A.Verbeke en W.Winkelmans (1998) *Sociaal-economische evaluatie van overheidsinvesteringen in transportinfrastructuur (Social-economic evaluation of government investment in transport infrastructure),* Garant Leuven/Apeldoorn

[11] ECMT (1998), *Efficient transport for Europe, Policies for internalisation of external costs*, European Conference of Ministers of Transport, OECD, Paris, in: Evaluatie van Infrastructuurprojecten, Deel 2, Capita Selecta, Onderzoeksprogramma Economische Effecten van Infrastructuur (OEII) (*OEII program of the Dutch Ministry of Transport*) Min. van Verkeer en Waterstaat en Economische Zaken, Den Haag,2000

[12] Jara-Diaz, S. (1986) *On the relation between users' benefits and the economic effects of transportation activities,* Journal of Regional Science, volume 26: 379-91

[13] SACTRA (1999) *Transport and the economy,* The Standing Advisory Committee on Trunk Road Assesment, Department of the Environment, Transport and the Regions, London. The paper by S. Jara-Diaz is found in Appendix C.