SIXTH FRAMEWORK PROGRAMME
PRIORITY 1.6.2
Sustainable Surface Transport

CATRIN
Cost Allocation of TRansport INfrastructure cost

D3 – Implications of cost recovery

December 2008

Authors:
Jan-Eric Nilsson, VTI
with contributions from
Magnus Arnek, VTI
Pedro Abrantes, ITS
Sofia Grahn-Voornevelt, VTI
Chris Nash, ITS
Jeremy Toner, ITS

Contract no.: 038422
Project Co-ordinator: VTI

Funded by the European Commission
Sixth Framework Programme

CATRIN Partner Organisations
VTI; University of Gdansk, ITS Leeds, DIW, Ecoplan, Manchester Metropolitan University, TUV
Vienna University of Technology, EIT University of Las Palmas; Swedish Maritime Administration,
University of Turku/Centre for Maritime Studies
CATRIN
FP6-038422
Cost Allocation of TRansport INfrastructure cost

This document should be referenced as:

Date: 2008-12-15
Version No: 0.9
Authors: as above.

PROJECT INFORMATION

Contract no: FP6 - 038422
Cost Allocation of TRansport INfrastructure cost
Website: www.catrin-eu.org

Commissioned by: Sixth Framework Programme Priority [Sustainable surface transport]
Call identifier: FP6-2005-TREN-4
Lead Partner: Statens Väg- och Transportforskningsinstitut (VTI)

Partners: VTI; University of Gdansk, ITS Leeds, DIW, Ecoplan, Manchester Metropolitan University, TUV Vienna University of Technology, EIT University of Las Palmas; Swedish Maritime Administration, University of Turku/Centre for Maritime Studies

DOCUMENT CONTROL INFORMATION

Status: Final submitted
Distribution: European Commission and Consortium Partners
Availability: Public on acceptance by EC
Filename: CATRIN D3 151208.doc
Quality assurance: Heike Link (DIW)
Co-ordinator’s review: Gunar Lindberg

Signed: Date:
Contents
D3 – Implications of cost recovery ......................................................................................................................... 1

1. Introduction and outline ................................................................................................................................. 5
   1.1 The assignment ......................................................................................................................................... 5
   1.2 The separate contributions ...................................................................................................................... 5
   2.3 Outline ................................................................................................................................................... 6

2. Market failures in the provision of infrastructure services. ........................................................................ 6
   2.1 Natural monopoly .................................................................................................................................. 7
   2.2 Public good ............................................................................................................................................. 8
   2.3 Externalities ......................................................................................................................................... 10
   2.4 Equity .................................................................................................................................................. 11

3. How is market failure dealt with? ................................................................................................................ 12
   3.1 Roads .................................................................................................................................................... 12
   3.2 Railways .............................................................................................................................................. 16
   3.3 Airports ............................................................................................................................................. 18
   3.4 Ports and other naval facilities ........................................................................................................... 19
   3.5 Demand growth and the climate .......................................................................................................... 20

4. Mechanisms to increase revenue generation .............................................................................................. 21
   4.1 Why may cost recovery be important? ................................................................................................ 21
   4.2 Congestion and scarcity charges ........................................................................................................ 23
   4.3 Distorting efficiency as little as possible .............................................................................................. 25

5. Clubs as a means to ascertain efficient infrastructure supply .......................................................................... 27
   5.1 Can clubs supply infrastructure services in an efficient way? ............................................................ 28
   5.2 Sharing the benefits of cooperation in parallel and serial networks .................................................. 30
   5.3 Is Eurovignette a Club? ....................................................................................................................... 33
   5.4 Sharing costs in Swedish road ownership associations ...................................................................... 35

6. Conclusions .................................................................................................................................................... 37

References .......................................................................................................................................................... 40

Appendix A. Aspects on marginal social cost pricing and cost recovery in the European transport sector .............. 42

1. Introduction .................................................................................................................................................. 42

2. Theoretical framework ............................................................................................................................... 43
   2.1 What characterise transport infrastructure? .......................................................................................... 43
   2.2 Efficiency aspects of infrastructure provision – the risk for markets failures and potential remedies ................................................................................................................................. 45
   2.3 Equity issues ......................................................................................................................................... 46
   2.4 Aspects on public finance .................................................................................................................... 50
   2.5 Aspects on cost recovery ..................................................................................................................... 51

3. Potential problems and inefficiencies in the European transport sector ..................................................... 54
   3.1 Roads .................................................................................................................................................... 60
   3.2 Railways .............................................................................................................................................. 66
   3.3 Airports ............................................................................................................................................. 70
   3.4 Ports .................................................................................................................................................... 74

4. Concluding remarks ................................................................................................................................... 77

Annex to Appendix A ....................................................................................................................................... 79

References to Appendix A .................................................................................................................................. 85

Appendix B. Congestion and Scarcity costs ........................................................................................................ 86
Appendix C. Cost recovery for transport infrastructure .......................................................... 93
Appendix D. Sharing profit in parallel and serial transport networks .................................. 107
Appendix E. Sharing costs in Swedish road ownership associations ................................. 135
1. Introduction and outline

1.1 The assignment

Under some circumstances, the marginal cost approach to infrastructure pricing leads to problems with cost recovery; a pricing policy which ensures that existing assets are efficiently used may not deliver revenue to pay for the costs for maintenance of existing, nor indeed for construction of new infrastructure. For different reasons governments may find this inappropriate and rather want to complement the marginal-cost-pricing principle with a requirement for a sector of the economy to break even. Efficiency would then be jeopardised.

The idea behind work package 2 is to establish the micro-aspects of requirements to recover costs over and above marginal costs. This is done for all modes of transport. More precisely, the objective of this report is to establish some core features of how each mode of transport is organised, to describe the implications of cost recovery requirements for each mode and to analyse different mechanisms which would ascertain that each mode covers a larger share of its own costs.

1.2 The separate contributions

The work package has sought to meet these objectives within the framework of five separate tasks. Each task has resulted in one or more reports which are provided in appendices. No formal referencing will be done to these appendices nor indeed to the wider literature referenced therein.

Task 2.1: Organisational structures and requirements. The purpose with this first task is to establish a frame of reference for the discussion of cost recovery within the separate modes of transport by clarifying the alternative organisational structures and cost recovery requirements used in different countries. This task is dealt with in a report by Magnus Arnek which addresses overall problems to establish efficient and equitable policies within the transport sector; the report is attached as Appendix A.

Task 2.2: Congestion and scarcity. This part of WP2 discusses how costs for congestion and scarcity can be charged for and the consequences of this for the cost recovery target. One purpose is to establish similarities and differences between transport and other industries with seasonal and daily demand variations and what this might mean for the transport industry. This issue is dealt with in a report by Pedro Abrantes and Chris Nash and is attached as Appendix B.

Task 2.3: Tradeoffs. The purpose here is to identify and discuss the basic tradeoffs involved in the choice between marginal cost pricing and full cost recovery. If industries with a high share of non-variable costs are subsidised, the taxpayer will have to foot the bill and the costs materialise in terms of dead-weight losses of the tax system. Under a no-subsidy policy, it is the users that will have to pay and the loss materialises as a sub-efficient use of existing facilities. These aspects of the choice of policy have been further developed and are discussed in several papers. Magnus Arnek addresses the overall tradeoffs from the perspective of a cross modal comparison in the Appendix A report. In addition, Chris Nash and Jeremy Toner reviews the instruments that can be used to generate revenue over and above marginal costs (in the first place Ramsey pricing and price discrimination schemes); this is attached as Appendix C.
Sofia Grahn-Voornevelt discusses alternative pricing and cost recovery strategies within the framework of multi-link infrastructure networks where links may be complementary to, or substitutes for each other; this is in Appendix D.

Task 2.4: Club approach: It has been suggested that a club approach could be one part of a policy to recover costs. The purpose of this fourth task is to establish the meaning of a club approach for each mode of transport and to discuss the applicability of dealing with cost recovery in this way. Within the framework task 2.4, a paper has been written by Sofia Grahn-Voornevelt on the sharing of cost and benefits within the framework of a club. In that paper, which is attached as Appendix E, a description is given of Sweden’s road associations, which handle and pay for a large share of the country’s tertiary road network, and the way in which they share the costs for doing so.

Task 2.5: Synthesis. The purpose of the final task is to make a synthesis of the first four tasks of the work package. The present text provides this synthesis. No further references will be given to the background documents in the appendices.

2.3 Outline
The rest of this task 2.5 report – i.e. of Deliverable 3 of Catrin – is organised in the following way. Section 2 delineates the major efficiency issues related to the provision of infrastructure services. Section 3 then describes our understanding of the financial and economic situation of different modes of transport across Europe. In this, we seek to make a distinction between a narrow and a broad definition of the cost recovery issue: while cost recovery in its broader sense includes also revenue from taxation of external effects (the concept is defined in the report) the narrow definition does not.

Section 4 reviews the different mechanisms at hand to deal with cost recovery issues in situations where marginal cost pricing will not suffice to recover the full costs of supplying infrastructure services. Section 5 discusses the merits of clubs for this purpose, as well as the broader issues at stake when people, firms, regions and/or countries could join forces in order to save on costs, increase revenue or improve welfare. The basis for this discussion is the concept of cooperative game theory. Section 6 concludes.

2. Market failures in the provision of infrastructure services.
The invisible hand of markets is widely held to provide for the efficient supply of a range of goods and services. This means that markets are able to meet the customers’ demand by investing in new facilities. Competition ascertains that profits are not supra normal so that customers don’t have to pay too much for their purchases. The competitive pressure also induces producers to manufacture at lowest possible costs.

It is, however, not reasonable to believe that infrastructure services can be made available for travel and transport by way of free competition between atomistic suppliers in the same way as for many other goods and services. The present section briefly summarises three reasons for why this is so; the natural monopoly aspects of infrastructure supply (section 2.1), its public good qualities (2.2) and the presence of negative externalities from travel and transport (2.3). Finally, section 2.4 have some comments on policy issues with an equity dimension.
Two dimensions of efficiency are in focus. The first is to ensure that existing resources – roads, railways, ports and airports – are used in an efficient way. It is well known that marginal cost pricing will achieve this objective: Charging for infrastructure use according to the incremental costs for admitting more vehicles, ships or airplanes to use them will make certain that there is neither too much nor too little traffic.

The second efficiency dimension concerns the addition of new production facilities – new infrastructure – when necessary. It is equally well known that this should take place whenever the expected revenue from charging future users plus the net non-monetary benefits exceeds the expected costs for construction and future maintenance.

A third efficiency concerns cost efficiency, i.e. the implementation of a certain task at lowest possible costs. Relatively less will, however, be said about cost efficiency in the present report.

2.1 Natural monopoly

A natural monopoly is at hand when it is cheaper to have one single firm supply all customers on a market rather than having several firms competing for demand, represented by D1 in Figure 1. The reason for why this is so is that average (AC) and therefore also marginal costs (MC) are falling within the pertinent range of demand. It costs a lot to have the original facility built, so the more customers, the lower is the cost per customer (the decreasing part of the AC curve). The addition of new users will moreover result in fairly low additional costs, which in the figure is manifested by a MC curve below the AC curve.

The presence of a single firm creates a welfare dilemma. A monopolist can be expected to charge a high price (for instance at pM in the figure). This will recover costs and also facilitate a sizeable profit. But the monopolist’s profit maximising policy is in conflict with efficiency since the high price will scare travellers away from using the facility; the infrastructure will not be used to capacity even though it would be cheap to admit additional vehicles.
It has already been indicated that an efficiency enhancing policy would call for a price at marginal cost ($p_{mc}$ in the figure). This would, however, not be sufficient to recover the full costs of supplying the service for a natural monopoly. In particular, revenue would not cover the costs for constructing new facilities when this is needed. This is at the core of the cost recovery issue: in a natural monopoly industry, a welfare enhancing policy does not guarantee that proceeds from (efficient) charging of infrastructure use are sufficient to cover the costs for (efficient) maintenance and expansion of infrastructure. To establish efficiency, it is therefore necessary to combine optimally low prices and an efficient investment policy with subsidies.

For the subsequent discussion it is important to emphasise that the cost recovery issue is related to the relationship between costs and demand. With the specific cost situation at hand in Figure 1, but with higher demand — for instance as depicted by $D_2$ — it may still be cost efficient to have one single firm supply the whole market but that firm could combine efficient pricing (at marginal cost) with cost recovery. Although not explicitly shown in the figure, this situation with high demand would call for an efficient pricing (a price set where $mc$ intersects the demand curve) which not only recovers all costs but which actually would generate a profit. We will return to this observation in section 4 below.

### 2.2 Public good

Two qualities distinguish a public from a private good; non-rivalry and non-excludability. The consumption of private goods result in rivalry since one customer’s purchase makes it impossible for another person to buy and consume the same good, for instance an ice cream. The consumption of a public good is not rivalrous in this sense; although one person listens to a
radio broadcast it is feasible for anyone else with a receiver to benefit from the same transmission.

The second distinction is related to the possibility to stop someone from consuming the respective goods. For private goods, excludability can be implemented through ownership; the seller keeps the good under lock until it is sold where after the customer can decide over whether to consume or to keep it (the ice cream in the fridge) for future sunny days. Once the radio broadcast is sent out it is, on the other hand, not feasible to exclude anyone from listening to it, meaning that it also has an excludability feature.

The presence of public good qualities in goods and services makes it difficult for a producer to ascertain full cost recovery. Since it is not possible to charge every user or indeed to stop anyone from listening once a show is on the air, there is a risk that this type of service never is produced. This is so even though consumers value the service at far higher levels than it would cost to produce then. The (potential) efficiency problem with public goods is therefore that of under supply.

As illustrated by the radio broadcasting example, there are several ways to make also goods with public good qualities available for consumption. The historical approach has in many countries been to have the public sector supply the market. This has been paid for over the tax bill or by way of license fees for all owners of receivers. Alternatively, and increasingly common, radio broadcasts are paid for by commercials, i.e. by private companies paying for having information about their products spread in parallel with the “core” programmes.

Infrastructure has some public good features. In situations with low use relative to capacity, additional users will not affect the possibility of existing vehicles to use the facilities. This non-rivalry aspect is obviously present during off peak, for instance during nights. In congested periods, infrastructure use is equally rivalrous as any private good.

It is relatively straightforward to charge for the use of ports, airports and railway infrastructure, meaning that these infrastructures are not a public good in the excludability sense. The reason is that relatively few vehicles make use of port facilities etc. so that simple charging schemes are technically easy and cheap to implement; a ship with poor records for paying the bills will not be admitted at the next arrival. For road infrastructure, exclusion has historically only been feasible by toll booths which is an administratively costly way to exclude those that don’t want to pay. It is, however, increasingly common that more sophisticated electronic devices are used to charge for use and to exclude or penalise those that don’t pay.

Moreover, fuel taxation could be seen as an indirect way to charge for infrastructure use. This is so since the consumption of fuel is proportional to the extent of driving, i.e. of consuming the service of the existing infrastructure. Infrastructure can therefore be seen as a public good first and foremost in periods with low demand when additional users don’t affect the possibility of existing vehicles to use the facility.

This quality happens to be very close to the presence of low marginal costs, discussed as part of the natural monopoly feature of infrastructure in section 2.1 above. For this reason, it is not reason to elaborate further on the public good features of infrastructure in the present report.
2.3 Externalities

The textbook definition is that an externality represents the (positive or negative) consequences of one person’s or firm’s consumption or production on the consumption or production of other persons or firms that the first party not necessarily takes into account. As a result, there may be too much (or too little with positive externalities) consumption or production of this type of commodity.

This long definition hides a range of features related to infrastructure use. A trip on a road by a vehicle imposes accident risks on other users. It gives rise to exhaust when burning fuel, it is noisy and it may cause some wear on the road. In congested situations, more vehicles will increase the travel time for all existing vehicles in the system.

The same line of reasoning also has a bearing on other modes of transport. There are some accident externalities in rail, air and naval services. Also ports, airports and railway lines may be used close to capacity meaning that there is congestion. Moreover, any use of transport infrastructure making use of fossil fuel generates greenhouse gases.

A functioning market is in principle able to internalise wear and tear and will handle congestion. This is so for private production of weekend relax services as well as air transport. Both holiday resorts and airlines have high fixed relative to marginal costs; it is expensive to build a hotel or to buy a new airplane, while adding guests or travellers is cheap as long as rooms or seats are available. Holiday resorts and airlines therefore charge less during off peak (perhaps somewhat more than marginal costs) than during the holiday season or when rooms and seats are expected to be fully occupied. The peak period guests therefore pay most of the bill for the original construction and purchase.

The same is, however, not automatically true of environmental and accident externalities, which do not spread via some sort of market. There is therefore a risk that transport is cheaper and consequently more extensive than what would be warranted from a social perspective.

Society has access to several means to cap consumption. One way to do so is to regulate – in the extreme to forbid – activities which give rise to (negative) externalities. The other mechanism is to impose a charge on the activity which generates the externality; the charge should then approximate the external costs. This is referred to as Pigou taxation. The traveller would then have to internalise the negative side effects and may as a consequence reduce this activity. This is indeed today the motive for the European Union’s official policy to handle externalities; see COM (2008).

An increasingly common alternative to regulation and pricing is the combined use of these two mechanisms by way of what is referred to as tradable emission permits. Only one aspect of this mechanism will be addressed here, namely its financial consequences. In the way that many of these cap-and-trade mechanisms have been implemented, and in contrast with a pricing mechanism, they will not generate any revenue for the treasury. It will be reason to come back to the implications of this for cost recovery.

Before doing so it should finally be noted that some side-effects of consumption and production activities are being internalised also by unregulated markets. This refers to the pecuniary consequences of a primary activity for the price of other activities or goods. For instance, a rise in the price of fuel can be expected to reduce the use of fuel. But it may also have consequences for the purchase of (new) vehicles and therefore for the vehicle prices. Another ex-
ample is that road improvements save time and vehicle operating costs for commuters. As a result, the value of their property may increase as a result of the better road. Since this side effect of an original change in the conditions in one market materialises through price changes in other markets, it is not an externality problem in the same way as most other side effects of travel and transport.

For the present report it is important to note that the Pigou taxes used to internalise external costs will have financial consequences for society; the surcharge is motivated as a tool for enhancing efficiency but it will also generate income for the treasury. It may therefore clarify the subsequent discussion to point to the difference between cost recovery from two different perspectives. Financial cost recovery has a bearing on the natural monopoly discussion, i.e. from the perspective of the corporation’s direct and indirect costs. Cost recovery from a social perspective should also include revenue generated by Pigou taxation.

2.4 Equity

From a social perspective, and even if the market was producing and allocating goods and services in an efficient way, the outcome may still be questioned if the allocation is not seen to be equitable. It is reason to sort out what this may mean when applied to the transport sector, and in particular to the issue of cost recovery.

Equity or fairness are two intertwined concepts with several dimensions. In the economic literature two stand out. A government program is considered horizontally equitable if similar individuals are treated similarly. The principle of vertical equity says that individuals who are in a position to pay more than others should then do so.

In many countries one political objective is that opportunities and standards of living for individuals and firms should be similar in all regions. This may be seen as a horizontal equity aspect. In the transport context it may be interpreted to mean that all travellers should have access to infrastructure of decent standard, irrespective of if there are many or few users. This means that more and better infrastructure than motivated by efficiency reasons should be built in regions where few inhabitants live. For the same reason, public transport in these regions should be subsidised.

A vertical equity dimension may be that governments sometimes charge fees to users in spite of that the marginal cost is low. This is in conflict with the efficiency arguments discussed in the context of natural monopolies and public goods. Examples include toll-roads, ports and airports. The equity argument may be that those who make use of the facility obviously benefit more than non-users. If users are believed to be better off than tax payers, they should – according to this argument – be paying for their consumption. In the extreme, it may be argued that a whole mode of transport should be covering its costs. For this reason, user fees are seen as an equitable way of raising revenue to finance public facilities.

In the literature, an extensive discussion addresses the trade-off between equity and efficiency, and there are two standard objections against the general use of equity objectives. The first and most general concern is that an equitable policy may mean that less will be available to allocate to those most needy than if an efficient use of resources could be implemented. Equity costs in terms of less becoming available to use.
The second concern, which has particular significance for the transport sector, is that equity issues, if important, should be dealt with by way of general economic tools such as general taxation and subsidies. It is, according to this line of argument, not cost efficient to adjust policies within a specific sub-sector of the economy in order to account for equity concerns. If inhabitants of a certain region should be treated favourably for equity reasons, this should be implemented in higher general allocations to that part of the country. Road investments should be forced to compete with other policies to improve the situation of people in the region.

There is also an equity argument between countries. Assume that it is efficient to charge the use of a certain piece of infrastructure – a road or railway tunnel, a port or an airport – below average costs. The citizens of that country would then have to foot the bill for that share of costs which is not paid for by users. If this infrastructure primarily is used for international traffic, this may be seen as unfair.

Irrespective of the arguments, it is a fact that policy making is affected by both equity and efficiency concerns. We will in section 4.1 get back to a discussion about why these concerns may be of relevance for cost recovery and for investment and pricing decisions within the transport sector.

3. How is market failure dealt with?

Infrastructure is of paramount importance to make economies function in a smooth way. Although not all infrastructures are provided directly by governments, a government has the ultimate responsibility that smooth and efficient transport is guaranteed. Against this background, the present section summarises our understanding of how governments in Europe in broad terms organises the supply of infrastructure services. The cost recovery implications of these policies are in focus of the review, and are dealt with in separate sections for the four modes of transport; road (3.1), rail (3.2), air (3.3) and sea (3.4). Section 3.5 finally adds a discussion about the overall growth of transport and the increasing interest in one of its market failures, i.e. the global warming effects emanating from the use of fossil fuels.

3.1 Roads

With some exceptions, road services are provided by the public sector. To be precise, the availability of roads is arranged by governments while governments have a choice between using in-house resources or to procure maintenance and construction from private firms. This may be seen as a way to handle the natural monopoly quality of road infrastructure supply. Since it is not cost efficient to have several roads competing for the same customers, the public sector provides the services in order to avoid high prices charged by a monopoly commercial operator.

A further common feature of road service provision is the institutionalised separation of on the one hand the charging of users and, on the other hand, the spending on investment in and maintenance of infrastructure. One ministry and its specialised road agency are responsible for taking care of existing roads and for the construction of new. The allocation of resources for handling this task is discussed in the government’s annual budget process. Another ministry, the Treasury, is responsible for charging by way of fuel taxation and indirect taxes on road users.
It is not obvious how this separation of spending and charging can be explained within the present analytical framework. It may, however, have something to do with the revenue generating potential of charging road users. UNITE (2005) established that the proceeds from taxation of fuel and ownership exceed the amounts spent on construction and maintenance in most (while not in all) European countries; these observations are reiterated in Appendix A. Road user taxation is therefore a potent mechanism for raising revenue for any government.

From the efficiency perspective, it may be reason to challenge these pricing and spending policies, and in particular the generation of a financial surplus. Taxes which (more than) recover costs are not compatible with the natural monopoly qualities of road infrastructure and the necessity to charge only at the (low) marginal cost level. Why should road users then pay so much?

One counter argument is that marginal costs are not always low. Parts of the road network are congested during parts of the day and week. Additional users will then imply high marginal costs. As illustrated by demand D2 in figure 1, high-demand periods may well warrant cost recovery charges.

It is, however, not obvious that the level of charging should be as high as it is today. In addition, the main charging mechanism – i.e. fuel taxation – is the same irrespective of if demand is high or low relative to existing capacity. A substantial efficiency problem is therefore the inability to differentiate charges according to the precise demand/capacity situation at hand. Congestion tolls in London and Stockholm only represent a first go towards a policy of more price differentiation.

The logic of the natural monopoly argument against present charging levels is therefore the following: parts of the road network is severely congested, some roads are congested sometimes while most roads are uncongested most of the time. Is it efficient to levy charges that more than cover the costs to build and maintain the road network, given this situation?

There are at least two arguments in favour of “high” taxation of road users; externalities and general taxation concerns. The externality issue arises since the natural monopoly discussion only addresses costs for building and maintaining the infrastructure. Adding external accident and environmental costs on top of these direct costs provides a complementary motive for expecting that an efficient policy would recover the full costs of road infrastructure provision.

The question is again whether existing fuel charges are appropriate to ensure efficiency when also external effects are accounted for. Much research has been done on this, and in very broad terms the following observations seem to be coming back:

- Passenger vehicles using rural, non-congested roads seem to be more than paying for their social costs, understood in its widest sense. Possibly excluding London and Stockholm, this is not so in cities where both health issues and congestion would require a higher level of taxation. Also main intercity arteries may be so congested so that higher taxes on passenger traffic may be warranted.
- On top of the arguments for passenger vehicles, heavy vehicles inflict substantial damages on the road when using them. Fuel taxation fails to account for the fact that the higher the weight per axle, the more substantial is the destructive power of a vehicle; this issue is specifically addressed in Catrin’s WP4. In addition, noise from heavy vehicles may be substantially more annoying than from private vehicles, adding to the
externality costs. Both level and possibly even more so the degree of differentiation across vehicle classes, therefore seem to indicate that heavy vehicles don’t pay their way with respect to the overall costs they inflict on society.

A caveat in this is that fuel taxes, which provide for the bulk of the price that road users pay, may not be the universally best way to internalise external costs. In particular, fuel taxation is a good way to internalise externalities which are proportional to fuel use, such as greenhouse gases, but less appropriate for local phenomena such as congestion. Moreover, it is not fit to handle the fact that wear of heavy trucks is related to axle weight, not fuel consumption.

The discussion of externalities must also account for the current surge of interest in the consequences for the global climate of using fossil fuels. If the current policy debate provides the scientifically correct conclusion, there still seems to be a long way to go before all externalities of road use – and indeed of all transport making use of combustion engines burning oil or coal – are internalised. We will briefly return to this issue in section 3.5 below.

The general theory of taxation and spending in the public sector provides a second argument in favour of revenue in excess of costs in the road sector. More precisely, public finance is related to the overall need to raise revenue for public sector activities which are not charged for along commercial principles. Defence, schooling and social welfare spending provide examples.

The general theory of public finance has landed in a fairly simple rule of thumb for the implementation of a policy which supports the efficient supply of any spending of this nature: Raise revenue where it hurts the least. The background is that any increase in the costs for providing a commodity, be it because of input prices that increase or because of a (higher) tax, will spill over in a higher price and therefore in a lower level of consumption. Since a tax does not reflect any real costs to society, this will reduce social welfare: taxation is costly since it distorts consumption and production patterns away from the no-tax, (presumably) efficient situation.

But commodities differ with respect to how much demand is reduced when the price increases. The recommendation is therefore to levy more taxes on goods and services with lower price sensitivity – lower price elasticity – than on commodities where the reaction to a price increase is swift and large.

Not least the taxation of vehicle ownership may be explained by this model. First, it is not vehicle ownership but vehicle use that produces externalities. Pigou taxes should therefore be levelled on usage rather than ownership. But, secondly, the taxation of vehicle ownership may display relatively low price elasticity, therefore providing the Treasury with an efficient source of revenue. This provides a complementary explanation to the use of road sector taxation as a source of general tax revenue.

This does not per se exclude the possibility to use vehicle taxation to correct for externalities, provided that these are not internalised directly, i.e. via the Pigou tax for instance of fuel. In this way, the vehicle taxation could function as a proxy for the preferred direct taxation and would then both generate revenue and provide information about externalities.

So far, the discussion has almost completely dealt with whether taxes and charges are at a level which ensures efficiency in the use of existing resources. Section 2 however also
stressed that efficiency calls for an appropriate level of investment in new infrastructure, and indeed for sufficient resources to maintain existing infrastructure. A complementary policy question is whether current policies deliver efficiency in these respects.

The standard way for markets to answer this question is to assess whether future financial proceeds generated by an investment would motivate the initial spending. No such tests can be performed in a sector with a monopoly service provider. A new road will improve the situation for users but – in particular given the application of homogenous charges over all roads – will fail to deliver more revenue than the existing road. It is therefore reasonable to expect that a monopolist following principles of profit maximisation would let the infrastructure deteriorate to fairly poor standards before being renovated and upgraded. The users after all lack alternatives under this strategy.

Rather than comparing financial revenue and costs, the accepted methodology is to apply social Cost Benefit Analysis (CBA) to assess the merits of investment in new, and maintenance of existing roads. This is a technique to account for consumer benefits which don’t materialise in the price of the commodity, in particular not on a monopoly market.

The question is therefore whether enough resources are allocated over the public budget to allow for spending on roads in an efficient way. The answer to the question would be important in that increases of, or savings in the road budget would reduce or increase, respectively, the financial net of the sector. Put in other words, today’s financial surplus would be smaller if more was being spent on construction and maintenance of road infrastructure, and vice versa.

It is, however, difficult to provide a straightforward answer. One important reason is that there are many competing uses of tax money. Since there seem to be few examples of CBA being used in sectors other than infrastructure, cross-sector comparisons are difficult to make. For the purpose of the present report, we therefore take costs as given and abstain from commenting on whether the optimal financial surplus would be higher or larger than it is today.

There are additional principles from the public finance theory policy which have bearing on taxation in the road sector. Diamond & Mirrlees (1971) for instance demonstrate a very general result saying that – excluding Pigou taxes levied to internalise external effects – intermediate products should not be taxed. For the transport sector this means that commercial vehicles should only be paying for the externalities they give rise to, but that they should not be requested to pay financially motivated charges. If the above interpretation of vehicle license fees is correct, these should be abandoned for efficiency reasons. We again abstain from probing further into this dimension of road sector taxation.

To sum up the arguments, it is possible to conceive of the road sector as a “production unit” which – in the same way as for many commercial “production units” such as holiday resorts – provides services which at the margin are cheap to produce. With this interpretation, the road sector is a decreasing cost industry.\(^1\) Current levels of charges, which generate revenue well above costs for building new and maintaining existing infrastructure, therefore seem to imply substantial inefficiencies, at least with respect to traffic on non-congested parts of the road network. Two arguments which weaken this concern have been discussed; the necessity of externality charging and the government’s use of transport as a general source of revenue.

\(^1\) In this, no account is taken of the externalities which may be related during the construction of new roads etc.
Both arguments imply that the observation that revenue from road users is well above costs for providing access to roads is not necessarily a signal of regulatory failure.

The objective of the present delivery is not to sort out whether road users are being charged “too much”. Rather, focus is on the issue of cost recovery. The conclusion for the present context is therefore that the issue of cost recovery is not of concern for the road sector as a whole. We will therefore not be pushing this discussion further in the present report.

3.2 Railways

During the introductory years of railways, and when the industry was an economically and financially vibrant part of many national economies, different railway companies were competing with each other. In contrast to roads, railways seem to have been vertically integrated with operations and infrastructure services provided as a single package. However, there may have been situations where different railways have been operating parts of their services on infrastructure owned by others.

By the middle of the 20th century, private railways had been merged into national, vertically integrated monopolies. Typically they were state owned. The process of consolidation was a result of an increasingly fierce competition from road transport. Shippers of commodities as well as travellers gradually got access to a reasonably cheap and convenient alternative. Railway operators could therefore no longer charge prices that would make it possible to recover not only costs for operating the services but also the high costs for infrastructure investment and for taking care of the network.

In the late 1980ties, Sweden was first to split infrastructure from operations. One bearing idea was the belief that the industry’s natural monopoly qualities primarily are embedded in infrastructure rather than in the operation of train services. Today, most European countries have done this exercise. In addition, the EU pursues a policy to allow for competition in the sector, to some extent on the tracks – i.e. with different operators competing head on. Many countries also implement a policy with bidding for the right to run a service under a monopoly franchise. In contrast to roads, the charging of track use is at least administered by the provider of the infrastructure services.

The poor financial results for the consolidated railway sector have however continued. In some countries, the industry seems to be required to cover its own financial costs. Figure 2 demonstrates that Poland, Hungary and the Baltic states have cost recovery ratios between 80 and 100 percent, meaning that the customers of rail services pay for train operating costs, for infrastructure maintenance and possibly also for more. The consequence of high charges is rapidly falling market shares for rail.

Based on the discussion in the previous section, this seems to provide a strong indication of a combined market and government failure. Marginal costs are low but infrastructure charging is so high so that passengers are induced to go by car and freight by truck. The railways receive insufficient allocations for maintenance and (re-) investment and tracks are underutilised. This is a particular problem if road users are not charged the full marginal costs. As discussed in Appendix A, some countries fail to charge even for roads’ direct financial costs, leaving much to be asked for when it comes to internalisation of external effects.
The financial situation in other countries (cf. the left part of figure 1) seems to be that train operations can cover their costs, sometimes with subsidies from different tiers of the public sector, while infrastructure still requires substantial subsidies, not least in order to spend on new, or to facilitate substantial upgrading of existing railway lines. Based on the natural monopoly logic, this seems to be a reasonable way to reduce the risk for market failure: subsidies guarantee that existing tracks are being used. In these countries, the failure of the market to provide the appropriate level of services is avoided. However, this does not necessarily mean that the current level of subsidy is optimal, an issue which is partly addressed by Catrin’s WP5 where new marginal costs are reported.

![Figure 2. Percent of Total Cost Covered by Infrastructure Charges in 2004.](image)

While the situation with persistent subsidies may be efficient, it has also prompted the questioning of cost recovery which is in focus in WP2. It is therefore reason to consider also the other arguments that may be at hand for and against (complete) cost recovery.

Railways emit noise which is a problem primarily in areas with high population density. In several other respects, the externality problem is less significant in the railway industry than in other modes of transport, at least for electrified services. This is in particular so since electricity generation is part of the Emission Trading Scheme, meaning that any externalities from using fossil fuels in the production of electricity is being internalised into the price railways pay for power.
Diesel locomotives may not be fully charged for their externalities. It is, however, reasonable to believe that even if they were, it would still be much cheaper per transport unit to use rail than any other mode of transport from this particular perspective.

Railways are, on the other hand, at least partly congested. Track capacity is not sufficient to meet demand on certain high-volume lines and during peak periods. This may not be a major problem as long as all traffic is carried out within and by one single company. The gradual opening up of the industry for entry will however make it increasingly pertinent to handle track scarcity by other means to construct the annual time table than administrative rules. The introduction of pricing instruments would obviously have consequences for the financial net of the industry, and will be further discussed in section 4.1 below.

The emerging road transport sector was one important reason for the railway industry’s gradual decline during the 20th century. Road transport remains the main alternative to the rail mode. Except for niche products – for example bulk freight trains and commuter trains in major conurbations – volumes are huge on roads as compared to railways. Even though competition within the railway sector still is fairly low, which would seem to facilitate monopoly pricing, it is not possible for a railway operator to price the services without due concern for the price of the competing modes. This goes both for freight but also passenger services where not only roads but also domestic flight competes fiercely for passengers with many inter city railway lines. The competition from other modes of transport caps the railway industry’s revenue earning potential.

Taken together, this seems to imply that there indeed is an efficiency motive for not discontinuing railway operations due to a poor financial result. The huge investment costs sunk in railway infrastructure can not be recovered by charging users. Line closure and indeed investment in new capacity has to be considered on a case by case basis by the use of the same analytical tools as in the road sector, i.e. Cost Benefit Analysis.

On top of this argument comes the possible inability to charge competing modes, in particular road use, for their full external costs. If externalities from trucks were fully internalised, rail would have been given a competitive advantage. In the absence of full pricing, there is a second best motive for subsidising the competing railways services. This would work to rebalance the relative prices of the competing modes; if one mode is charged too little, it could be reason to balance this by charging the competing mode less and by investing less-than-optimal in the undercharged mode and more-than-optimal in the competing mode (cf. Nilsson 1992). First best optimal would of course be to charge each mode for its social marginal costs.

### 3.3 Airports

Most airports are operating with a degree of natural monopoly: it is not viable to have several adjacent airports compete head on with each other. To a degree, this conventional wisdom seems to be challenged by the growth of low cost airlines. One part of their strategy is to operate from non-hub airports with modest landing fees and little congestion. At least in some countries, major airports are therefore challenged by fringe competitors.

This does not stop the major airports from being virtual monopolies. In many countries, the downside of this is dealt with by having the airport within the public sector, the underlying idea being that this is a means for ascertaining that prices don’t reach monopoly levels. Else-
where, in particular in England, airports have been privatised but are then subject to regulation in order to prevent them from acting as monopolists.

There also seems to be a degree of cross subsidisation within the industry. In some countries, parts of the revenue from hubs are dedicated to make up for losses at secondary airports. Based on a belief that an airport is a vital tool to attract business, secondary airports may also be subsidised by the local communities where they are located.

Given these caveats, airports seem to make a decent living. The revenue from take-off and landing charges and charges from handling passengers, luggage and freight, in combination with income from parking and vending licences, means that the industry is not a financial burden for the society at large.

In the same way as for roads, there still is an issue with respect to degree of internalisation of the external costs for air traffic. This has repercussions for the competitive situation for competing modes. The financial implications of full internalisation of externalities are, however, unclear. One reason is that much of air transport is undertaken over the air space of other countries. It is therefore not obvious how a complete internalisation would affect revenues. This is even more so to the extent that externalities are handled by way of cap-and-trade instruments.

### 3.4 Ports and other naval facilities

For centuries, ports have been economic units of major importance in international trade. The important role of ports has not decreased. Today nearly 90 percent of goods exchanged in the international trade process in the world rely on maritime transport. On average, sea transport is currently growing at an annual rate of three percent. This development means that well functioning ports are important for governments, in Europe as well as in other parts of the world.

Ports offer infrastructure services, vessel related services and cargo related services. The respective activities are often carried out by different firms which are coordinated by the port authority. Three different organizational models are used.

- The landlord port where the port authority owns and manages the port infrastructure. Private firms provide port services and own their assets – the superstructure (buildings etc.) and equipment (cranes).
- The tool port where the authority owns both infrastructure, superstructure and equipment. Private firms provide services and rent port assets through concessions or licences.
- If a port authority is responsible for the port as whole this is referred to as the services port.

Countries that have adopted an Anglo-Saxon approach to charging for port services have a clear commercial orientation in which users (mainly shippers) bear all costs generated in the production of various services provided by the ports. According to this doctrine, the port should be profitable or at least generate revenue to cover its costs so that no tax subsidies are required. The continental approach to charging is less commercially orientated. Instead, ports are considered as being part of the society’s overall infrastructure, much like roads and railways. Ports are seen to be vital in contributing to economic and regional development and the
government is therefore willing to promote ports in fulfilling this role. Public subsidies often make up for substantial parts of the costs.

The cost structure of a port is like most other infrastructure; it costs a lot to have it built and equipped with cranes etc. Up towards the capacity limit, the additional expenses for admitting vessels are low, but at some level of demand interactions between the different users start to induce disturbances. The scale economies are even higher when it comes to navigation aides, where there are virtually no marginal costs to admit additional users.

The situation with respect to under pricing of emissions from naval vessels seems to be the same as for airlines. In the same way as for airports, competition between cities creates a check on the ability to utilise natural monopoly powers. Market power therefore seems to be of secondary importance only.

There are, however, situations where ports are subsidised in order to stop services from being discontinued, which then may be in line with the natural monopoly qualities of the services. This does not seem to result in a general outcry for national subsidies, but the support grows out of concern over employment opportunities etc at the local level. Another observation is that there seems to be no outside pressure to increase the cost recovery ratio of ports as an aggregate.

Overall we therefore conclude that it is not reason to be concerned about cost recovery issues in the maritime sector. The industry seems to be thriving and able to pay whichever charges they are asked to pay. There is competition between ports which has two consequences; to cap monopoly pricing and to induce local communities to provide subsidies to stop their own ports to loose all services.

3.5 Demand growth and the climate

There are, finally, two issues which warrant some additional attention from the perspective of the transport sector at large. The first is with respect to overall demand for transport. The globalisation of industrial production, meaning that countries increasingly specialise in fields where they have absolute or relative comparative advantage, means that products have to be transported to an ever increasing degree. As a complement to the growth of trade, the number of business trips increase.

On top of this, leisure travel – both domestic and international – is increasing. This is being supported by a long-lasting increase in income levels as well as increasingly efficient and cheap travel services. The deregulation of international air traffic and the growth of low-cost airlines is one example. These trends have been complementary to, and possibly at least partly triggered by, a general reduction of travel costs. The gradual upgrading of infrastructure has been an important aspect behind this trend.

Railway transport has benefitted less than the other modes from this overall growth in demand for travel and transport. One reason is that rail first and foremost has its cost advantage in large-volume bulk transport and commuter services within and between major conurbations. It is first and foremost in these types of relations that the high costs for infrastructure construction and upkeep can be spread across a large enough number of customers to make prices competitive. Another reason may be the inability of nationalised monopolies to respond to rapidly transforming market conditions.
Another reason for the rail industry’s decline is the inability to charge for the full social costs in other modes. As discussed in section 3.2, the gradual shift from rail to other modes may have been excessive if users are not paying the full external costs in competing modes. This provides a further background for the present paper’s most basic argument, namely that there are strong arguments in favour of persistent subsidies to the rail industry. This conclusion does, however, by itself not have any bearing on whether or not current subsidies are at an appropriate level.

When this is written, the world market price of crude oil hits new record levels almost by the day. This will affect the relative price of different modes of transport and shift demand in favour of rail. The degree of this shift is, of course, difficult to foresee.

This goes in parallel with the other major policy issue of our time, namely the possible global warming consequences of burning fossil fuels. Again, the policy for dealing with these externalities will affect the relative price of transport to the benefit of rail, and – again – it is difficult to see how large the effect will be.

It is well known that the most potent and cost effective policy instrument for internalisation of greenhouse gases is to charge the use of fossil fuels higher than at current levels. Except for affecting the relative price of different modes of transport, this will also generate revenue for treasuries around the world. If some other policy instrument – such as emissions trading – would be used the relative price of modes will still change, but the consequences for ministers of finance will be less significant.

The overall conclusion from this line of reasoning is that the arguments for financial support to the railway industry seem to be substantial also in a long run perspective. It is strong reason to believe that it is efficient to use railways also in future. To make rail services competitive, subsidies are required since prices to recover costs would scare away users. This further supports the necessity to deal with the financial aspects of this policy, which indeed is the task of WP2.

4. Mechanisms to increase revenue generation

In focus of the present paper are those parts of the transport sector where efficient charges are levied but where these are not sufficient to fully recover the financial costs of the activities. Section 3 has demonstrated that the prime reason for concern in this respect relates to the railway sector: In the absence of subsidies, there is a clear risk for that supply of railway services would be well below efficient levels.

While the arguments in favour of charges below average costs in the rail industry are strong, it is reason to further dissect the possible motives for this may still be a problem. Section 4.1 therefore dissects the possible motives for (full) cost recovery. The rest of this section then deals with mechanisms that would increase cost recovery. Section 4.2 focuses the possibility to introduce charges which would contribute to both efficiency and cost recovery, i.e. fees to handle scarcity and congestion. Section 4.3 then discusses the two classes of tools that over the years have been suggested in order to generate revenue which exceeds those from marginal cost based prices. The point of departure for that discussion is to levy such charges in order to inflict as little harm on efficiency as possible.
4.1 Why may cost recovery be important?

The starting point for this discussion is that efficient charges are not sufficient to cover all costs for an activity, but that activity is still considered socially valuable. General tax revenue is therefore required to keep the services running or to keep a branch line open for traffic. It has been argued that the railway industry first and foremost is the subject of concern.

There are however also efficiency motives to back the cost recovery requirement and in addition possibly also an overall political restriction. The basic motive for cost recovery is related to the incentives created when an activity is required to recover its full costs in comparison to when this is not necessary. When “someone else” will pay at least part of the expenses for operating for instance a branch line of the railway network, the pressure to keep costs down is reduced. In addition, the organisation is given reason to develop skills to argue for more (external) resources, i.e. to expand activities compared to what would have been relevant under a cost recovery constraint. The presence of a soft budget constraint will therefore have repercussions for the cost efficiency of the activities: costs will be higher under a subsidy than under a stand-alone regime.

But (re-) introducing a cost recovery constraint would come at the price of reduced allocative efficiency: Existing facilities would be used below capacity. The argument in this paper is therefore that it may promote overall efficiency to accept that traffic is subsidised in order to stop it from being discontinued.

This argument however points to the necessity to trade off cost efficiency and allocative efficiency. The presence of two partly conflicting efficiency targets means that the industry should be required to recover somewhat more than marginal costs in order to reduce the most severe consequences of a soft budget constraint. Subsidies should moreover be designed in order to preserve at least a degree of cost saving incentives. This can, for instance, be achieved by requiring the industry to pay for at least a share of the costs for new investments.

This line of thinking can be taken one step further. To repeat, the argument against cost recovery is that it would negatively affect allocative efficiency. The negative effects for the subsidised industry should, however, be balanced against the detrimental efficiency consequences of general taxation. As indicated in section 3.1, any tax distorts behaviour and introduces a dead weight loss for society. The conclusion would therefore be that the efficient level of subsidy should be settled in order to balance the allocative distortions from general taxation on the one hand and charges above marginal costs on the other.

Although there are efficiency arguments in favour of subsidies to natural monopoly activities, charges should be set above marginal while not necessarily at average costs. There may in addition be other political motives to require an industry to cover its own costs. In particular, it could be seen as an equity issue – an issue of creating a level playing field – that all activities in general and all modes of transport in particular, including railways, are able to cover their own costs.

While the previous discussion suggests that the overall arguments for full cost recovery are weak, it is still commonplace to see this line of arguing in the policy debate. Moreover, there is competition for public funds in most countries. This may induce decision-makers to require

---

2 To the extent that competitive tendering is used as a mechanism for subsidising services, the soft budget constraint issue is wholly or partially handled.
cost recovery as a means for softening the pressure on them for more funds. Irrespective of the precise motives, the rest of this report therefore takes the necessity to raise revenue above marginal costs as a datum.

4.2 Congestion and scarcity charges

In road networks, congestion is defined to be the additional time inflicted on all existing vehicles in a system which is used close to capacity when an additional vehicle enters into the network. Congestion costs are born by the society of road users. From the perspective of the marginal driver, congestion is, however, external since there is no mechanism to ascertain that this driver takes the increased costs for existing drivers into account before entering the system.

Since railway services are planned and scheduled long before a train leaves its departure station, it is however reason to make a distinction between congestion and scarcity in the railway system. A first aspect of this difference is that if a certain train is the direct cause of delay, for instance due to mechanical problems, the costs for others should be covered by a performance regime.

Congestion costs on a railway line used close to capacity are, however, incurred when running an additional service causes additional delays to other services. Even when the additional service does not directly cause delays, unexpected delays to one service are likely to lead to additional delays to other trains as there is little slack in the timetable to allow the system to recover. It is, in other words, more difficult to get back to the original time table after that a disturbance has occurred the more crowded is the system. This is the (external) congestion cost in the rail system.

Scarcity costs rather appear during the planning of future railway services. Scarcity is at hand when the operation of one train service prevents another from operating, or requires other trains to use inferior paths. Since a network could be used to run many possible combinations of services, it is difficult to uniquely determine its capacity and therefore the opportunity cost of a given service. Furthermore, the impact of an additional train of a particular type on the paths available to other trains can differ significantly according to the precise mix of traffic on the line. At the same time, the value of a standard slot to a given operator will also vary in time and space, and will depend on what else is running on the network.

Both congestion and scarcity are external effects in the way the concept is defined here. These externalities would, however, not create efficiency problems if they were incurred in a monopoly network. When one and the same owner runs all types of services, that firm will itself have to bear the consequences of all delays and scarcities; these effects are therefore automatically accounted for, i.e. they are internalised. This is not so if the railway network is used by independent operators.

Where the number of road users is large, congestion charging is the obvious solution to internalisation. Since the number of railway operators is not so large, it is obvious that administrative rules for prioritising services, both when time tables are planned and when delays occur, are an option. This is indeed also the way in which these problems historically have been and still are handled, meaning that the railway industry’s scarcity problems are resolved during the time tabling process well before services commence.
In the same way as for any administrative procedure, the problem in time tabling is the difficulty for the planner to know which service is more valuable than the other and therefore should be given priority. This is, on the other hand, where the prime benefits of the price system lays, i.e. to let operators react to posted prices which encourage some from adjusting their demand more than others.

A first approach for internalisation of scarcity in a railway network with different service operators competing for access would therefore be for the infrastructure body to post prices as the initiating step of the time tabling process. For lines with high demand, it would then be more costly to let a train run during peak than off peak periods. Knowing about these costs, operators are given reason to think through their service plans before demand for slots is submitted. This corresponds to the way in which congestion charging works in Stockholm and London, where some drivers leave their car to use other modes of transport, change the timing of their journey, etc.

When the reaction to congestion charges for road use is estimated, analysts can rely on behavioural patterns of a very large number of users. It is less straightforward to estimate a priori precisely by how much, for which parts of the network and during which hours that train operators should be charged. As a result, the original rail charges could be too high (resulting in that more traffic than necessary is discontinued) or too low (with scarcity situations still to be dealt with). This may call for several iterations of the time tabling process before the scarcity problem could be solved.

An alternative would be to turn the responsibility for calculating a price upside down: The operators submit their demand for departures in combination with a price which they would be willing to pay in order to be given the right to run a service. Priority is then given to operators with a high value-of-service. This is in reality a bidding process or an auction, which may also require several iterations before an equilibrium can be established.

There are many practical and principal issues that would have to be solved before a system with posted prices or with auctioning could be implemented. The appropriate way to solve the scarcity or congestion problems in the railway sector is, however, not in focus here. The reasoning however points to important conclusions for cost recovery in the railway industry: if demand for railway services is high relative to available capacity, and as long as marginal cost pricing does not include any component to account for scarcity or congestion, there is reason to increase the levy on railway operators.

Higher prices would indeed increase the financial pressure on an industry which is already bleeding. In the aggregate, a functioning pricing mechanism may still improve the efficiency of the railway system as a whole. For instance, operators are given reason to consider ways to bypass the charges, for instance by rescheduling (freight) trains to periods with excess capacity, by terminating long distance trains at commuter stations at the outskirts of major conurbations or by merging two trains into one single long train. In particular, when demand for railway services keeps growing, it will give time for expanding capacity. And it will help infrastructure planners to identify where in the network that additional capacity is most badly needed.

Congestion and scarcity charging, both using posted prices and iterative bidding, would moreover contribute to the funding of the infrastructure. Moreover, it does so in an efficient way. It is reason to recall the parallel between a railway and the holiday resort which recovers
its fixed costs during peak periods. Even though railways might not be able to use peak load pricing to generate full cost recovery, any extra incomes generated in this way will actually increase efficiency.

Parts of the financing problems of railways are therefore due to the inability to charge for congestion and scarcity. There are therefore extremely strong arguments in favour of the development of techniques to handle these issues in a deregulated rail sector.

4.3 Distorting efficiency as little as possible

The perspective in this section is that of an organisation which has two sources of revenue; from charging its customers and from state subsidies. At the outset, the organisation’s charging revenues are based on a policy of marginal cost pricing. However, for one reason or other the government wants to reduce its subsidies. To balance this shortfall of revenue, the organisation is instructed to increase charges, but to do so with as small consequences for efficiency as possible. There are two main approaches for doing so, Ramsey pricing and multi-part tariffs.

**Ramsey Pricing:** The logic given above for levying taxes so that dead weight losses are minimised carries over also to Ramsey pricing of a firm’s products. The basic rule is to charge a larger mark-up on goods and services which have a relatively lower elasticity of demand and to up the price less in the relatively elastic segments of the market. This is sometimes said to be a way “to charge what the market can bear”, in so far as the seller tries to ask for more from those customers that still will purchase.

Ramsey pricing is also referred to as second best pricing, the reason being that the financial outcome of first best – pricing at marginal cost – for one reason or another is not believed to be sufficient. What comes second is then a policy which satisfies the restriction (break even or some specific revenue target) at lowest possible sacrifice of efficiency.

Applied for infrastructure use, the implication of the Ramsey principle means that the service provider has to assess the business surplus of the different train services and to increase the charge more where the operators make a surplus and where they don’t have access to an alternative way to run the service. This in turn may point to services where travellers or freight customers in their turn are captive for the rail operator, i.e. where the price of using alternative routes or modes are substantial, and where no substitution will take place.

If two train services have the same price elasticity of demand, the Ramsey rule generalises into also accounting for the share which the fee takes in total costs for the operator: for equal price elasticity at the final market, the smaller is the share of access fees in the total cost of an operator, the higher should the mark up be, all other things equal. The logic for this goes back to the impact that the mark-up would have on the price of the final product. If, for example, the price elasticity of demand for train trips from London to Brussels were the same as the elasticity of demand between London and Marseille, then the mark up which Eurotunnel charges should be proportionately bigger for the Marseille than for the Brussels train, assuming it costs more to run a train from London to Marseille than it does to run one from London to Brussels.
The Ramsey rule can also be generalised to situations with interdependent demand. The infrastructure agency runs a network that has a remit to break even and where different operators, or different types of services run by the same operator, make use of the respective branch-lines. There may in addition be a degree of cross elasticity between these different services. Due to an interchange of passengers at station B, the demand for services between B and C may partly be explained by the price of the A to B trip.

When there is price independence between services – in technical terms when the cross-elasticities between the products are not zero – the Ramsey rule can be generalised: Price should now exceed the marginal cost for letting a service use the infrastructure by a greater amount:

1. the tighter the budget constraint;
2. the lower the own-price elasticity;
3. the higher the cross-elasticities (the greater the substitutability between i and the firm's other products).

**Multipart pricing**: Ramsey pricing seeks to charge different prices to different (sub-) markets which, inter alia, can be identified by different price elasticities of demand. In contrast, multipart or non-uniform pricing seeks to charge different prices to consumers in the same (sub-) market. In the simplest version of a multipart tariff – the two-part tariff – the customer is required to pay a fixed fee to be given the right to consume the commodity. The consumption price may then be equal to the marginal production costs, meaning that consumption at the margin is not affected by the introduction of a fixed charge.

In the utility sector, a two part tariff means that the customer pays per each kW of electricity. In addition, a fixed annual or monthly payment is made which does not relate to how much that is consumed. The average price per unit will then fall with consumption.

In the ideal case, the fixed fee will not affect the volume consumed. The tariff is then merely a way of having the customer give away more of the consumer surplus compared to the standard situation with one single price.

There is, however, always a risk that the fixed fee will scare away some buyers and the two part tariff will then distort efficiency. To reduce this risk, the supplier may offer customers a multi-part tariff. In its simplest form, there is then a choice between a single price – which then presumably is higher than the marginal cost – and a two part tariff. The two-part tariff may also generalise into a structure with more than one fixed component.

There are several ways to apply multi-part tariffs to the use of infrastructure supply. One example could be to require each operator to pay an annual fixed charge in order to be allowed into the system. Under a system with free entry, this design illustrates the competitive hazards of such systems, i.e. the risk that the fixed charge will constitute a high step for entry into the industry. Competition would thus be jeopardised even though there is no real cost which should motivate that, entry by new service providers is staved off. Under a system with monopoly franchising, this would not be an issue.

Another type of fixed charge would be to ask one or a group of users to jointly pay the costs for construction of a new, or upgrading of an existing line. This could be the full costs for a
project or some predetermined share thereof. If the payment is upfront, it is very much like a Public Private Partnership (PPP) where the club members form a Special Purpose Vehicle, while an annual fixed payment is more of a multi-part tariff nature. A particular form of a PPP is the club solution which is discussed in the next section.

The Arlanda airport rail link illustrates the PPP type of application. A private consortium built a branch-line and a station beneath the airport. These costs are to be recovered from user charges – i.e. ticket revenue from the Arlanda Express train – over a 45 year period. The agreement between the government and the private consortium also gave the latter control over access to the new piece of infrastructure. If other service providers want to make use of the assets, they have to negotiate a charge which is well above the (marginal) costs for letting them use the submerged station under the airport. This illustrates the potential risk for allocative inefficiency, i.e. that the levy will deter entry and result in underutilised infrastructure. Cf. further Nilsson et al (2008).

An alternative to charging for entry into the system is to combine a usage charge with a levy on vehicles; this is the model used in the road sector, where annual license fees generate substantial revenue for the Treasury. Applied to the railways sector, the obvious risk is that the vehicle fleet would be downsized, in spite of that there are no actual social costs that are saved by selling off old or not buying new rolling stock.

**Conclusions:** The important conclusion from this discussion is that there are ways to overcome financial restrictions in ways which are less harmful than others. There are, however, no guarantees for that prices above marginal costs are not harmful for efficiency. While some prices above marginal costs may have very little impact on the consumption pattern, this is typically not so. As a consequence, establishing financial restrictions on the overall welfare objective should always be a delicate matter of balancing up- from downsides of the extra charges which are levied.

### 5. Clubs as a means to ascertain efficient infrastructure supply

Club theory sat out to fill the gap between theories on private and pure public goods. For private goods consumption rivalry is complete and exclusion is costless, while for public goods consumption is nonrivalrous and exclusion is infeasible. A club good is defined to be between these extremes, “close” to a public good. The theory was developed by Buchanan (1965), who viewed clubs as a private nongovernmental alternative to the optimal provision of a class of public goods, later known as club goods.

Club theory may be of relevance for infrastructure supply at large and for the issue of cost recovery in two ways. First, the establishment of a club could be a way to ascertain an efficient supply and pricing of infrastructure without having to get backing from the Treasury. Secondly, the theory could provide a conceptual model for better understand the incentives inherent in different institutional models for infrastructure supply.

The overall qualities of club theory are further described in section 5.1. It is argued that there may be situations where the establishment of a club could replace the public sector in providing club services. The rest of this section then reviews some possible transport sector applications of club theory, namely the voluntary sharing of costs and/or benefits within a group of beneficiaries. Section 5.2 discusses the joint pricing of road or railway links which are either
complements to, or substitutes for each other; section 5.3 reviews Sweden’s road ownership associations; section 5.4 discusses whether the Eurovignette system could be thought of as a club.

5.1 Can clubs supply infrastructure services in an efficient way?

The following features distinguish club goods from pure public goods (see also Sandler and Tschirhart 1997). First, clubs are voluntary; members choose to belong to the club because they anticipate a net benefit. As a consequence, the net gain in utility associated with membership exceeds or is at least equal to the membership fee. This is not necessarily the case for a public good since such goods may harm some recipients who don’t appreciate having to consume it. An illustration is provided by the pacifists who don’t approve of defence protection (which then is a pure public good).

Second, in contrast to a pure public good, club goods involve sharing that may result in congestion or crowding. Crowding depends on some measure of utilisation; the number of members, the total number of visits on the club’s premises or the ratio of members’ utilisation to the number of units provided. Crowding depends positively on a measure of utilisation and negatively on the provision level. The provision of club goods therefore involves two allocative choices in contrast to the single provision choice of pure public goods: Membership size must be established along with the provision level of the shared good. Since the number of members affects the optimal size of the club good, and vice versa, these decisions must be made simultaneously.

Third, club goods require an exclusive group of members to be established. In contrast, pure public goods are associated with inclusive groups, since additional users can bring down per-person fees and impose no crowding costs on others. At the margin, the cost-sharing gains from admitting additional members must be balanced against the losses associated with crowding. This also, fourth, means that a club requires an exclusion mechanism so that members can be charged tolls and non-members kept out.

When the club decisions are represented as a cooperative action, the resulting outcome will be a Pareto optimum for the members. This is so since members join the club and pay their dues because of the perceived net benefit of membership.

The establishment of a club does not necessarily represent a Pareto improvement for the economy as a whole. The reason is that the members may decide to exclude applicants once a club has been established and the original investment into “the club house” is sunk. This would not be of concern if another club could be established by those excluded from the first. If so, the number of clubs, and the services provided by these clubs could multiply until all applicants are admitted somewhere.

Since transport systems usually are subject to crowding and since exclusion is possible, they have several features in common with a club good. Let us therefore toy with the possibility to let a club provide a country’s road infrastructure. Membership would be voluntary, constituted by the purchase of a vehicle and the payment of a vehicle registration fee. Charges for usage would – in the same way as today – be made part of the fuel price.

The gain from letting users provide the transport system via a club is that the planning and charging decision becomes more directly in line with the interest of the users. It is, however,
difficult to recognize how a vote on the optimal supply and charging of usage would provide for better decisions in these respects than today’s system where decisions are taken by national parliaments. The benefits of direct democracy within a club clearly vanish when the number of users gets too large.

Germany’s heavy vehicle duties seem to have been assessed with reference to a club solution. (Parts of) the costs for motorway provision and maintenance have been allocated across weight classes along these lines. The reference to the club concept to allocate costs across user groups is however not relevant due to the voluntariness of the club concept. There is no way for a vehicle owner to opt out of the club if he or she does not approve of the majority’s decisions.

There are, however, other partitions of road users where the club concept may be of some relevance. Section 5.3 below argues that this may be so for Sweden’s road associations. Moreover, section 5.4 discusses the Eurovignette from the perspective of club theory.

A club to organise the provision of railway infrastructure would have the operators of railway services as the constituting members. Compared to roads, this may under some circumstances provide for a handy number of members, with a few up to perhaps thirty operators. As long as members would not be too numerous, it would obviously be feasible to take votes on network expansion and on the sharing of costs between participants in a way which is far closer to the interests of members than when a parliament votes on these things or when it is handled by a state bureaucracy. Moreover, club ownership of the facilities could possibly make it feasible to undertake maintenance activities in a more cost efficient way than when infrastructure is operated by a public sector agency.

From the perspective of society, it is conceivable that some sort of deal could be worked out between the club and representatives for the government to handle the overall issues raised by the natural monopoly features of the railway system. The club could be subsidised in order to ascertain that the railway’s full potential is realised, but club members would handle a larger share of costs for infrastructure provision than at present, possibly at lower costs.

From the national perspective, the shortcomings of this club lie in its incentives with respect to membership and network renewal. A club would be reluctant to admit new members, even if track capacity would be sufficient to accommodate the operation of more trains. The reason is that an entrant would take away customers from the incumbent(s) running the specific service. Alternatively, an existing member may be interested to exploit a new business concept originating from an applicant to the club. The club would therefore not make the market dynamic.

The same logic may apply to investment projects. This is so if the existent shortage of capacity is used by operators to charge users a high price; this would no longer be feasible if and when capacity was expanded. In situations of this type, a club would have incentives to delay projects that are warranted from an overall welfare perspective.

A similar line of argument carries over also to the club ownership of airports and seaports. It is clearly feasible to organise an airport as a club made up of the operators flying to and from it. The airport club would, however, afford a degree of local monopoly to its owners. In spite of spare capacity, they may therefore have incentives not to admit new members – new airlines – since that would undermine their possibility to exercise their market control. In such
situations, club solutions – although efficient from the perspective of its members – may not be so from the national perspective.

In principle, it would be possible for a national government to be involved in the establishment of a club in order to cap these incentives. The club could, for instance, be obliged to establish rules for the acceptance of new members in order to reduce the risk for ossification of the market structure. The problem is, however, that such rules would also have to account for the costs that club members have sunk in the construction and maintenance of the assets controlled by the club. If not, potential entrants could postpone their membership application until after that for instance the costs for a substantial capacity expansion has been sunk. The problems in doing this are similar to the wish of fringe telecom operators to be given the right to roam on the incumbents’ (sunk) investments in masts and other equipment.

The creation of clubs is obviously no panacea to establish efficient supply and usage of infrastructure, or indeed to handle the deficits in decreasing cost activities. The club concept may, however, in specific situations contribute to the understanding of these classes of decisions. We now revert to discussing some examples of this.

### 5.2 Sharing the benefits of cooperation in parallel and serial networks

This whole section should discuss and explain more the outcomes of the analysis. Too often it is only phrased as “the analysis demonstrates” without demonstrating. This should be brief but a bit more informative!

Europe’s road and railway transport networks are made up of sub-networks, each comprising links which are under the jurisdiction of different countries. The pricing schemes for using these networks may differ significantly. Since the networks are connected to each other, the charging regime in one country has consequences for demand and presumably (optimal) pricing in other countries. It is then natural to consider the benefits of cooperation between countries when it comes to pricing decisions and in particular how to handle the benefits from such cooperation. If European countries are able to cooperate on their pricing policies, they can be seen as creating a club in the sense discussed in section 5.1.

Previous research has addressed the issue of optimal pricing in transport networks made up of links in separate countries. The models assume the owner of each link to toll the traffic using the link, and that the infrastructure manager of each country maximises a welfare function with respect to the toll. Importantly, the infrastructure manager only considers the welfare of his/her own country in this optimisation.

The work undertaken under the aegis of Catrin WP2 and reported in Appendix D, complements previous research by looking deeper into the incentives for cooperative behaviour among countries that are part of a network. In particular, mechanisms for increasing the chances of cooperative behaviour among the owners – i.e. to establish a club – are developed. The analysis does however not address the non-cooperative process which leads to the establishment of a deal, and in particular not whether the club would have all link owners within its ranks; one or more single countries could choose to opt out, the club still being established by the remaining countries. The focus is on increasing the chances for a successful transfer from a situation without cooperation to a club.

Two types of networks with congestion are considered:
• A parallel network where a number of links are substitutes to each other, meaning that transit traffic can choose between alternative routes to get from origin to destination. This could be alternative routes to travel or transport by road or rail between northern and southern Europe through Germany, Austria and/or Switzerland.

• A serial network with a number of consecutive links, together forming a transport corridor. Transport and travel from the Nordic or Baltic countries to France would have to make use of links in Poland (for the Baltic States), Germany, Belgium, Holland and France.

In the parallel case, the links are substitutes. As expected, the analysis demonstrates that cooperation makes it possible for owners to jointly operate as a monopoly. The establishment of a club will therefore result in higher revenue for the owners of the networks at the expense of transit traffic that would have to pay higher tolls. However, the larger the number of alternative routes, the lower would the tolls be and the higher is the welfare for society at large.

In the serial case it is demonstrated that cooperation is not only beneficial for the owners of the network but also for the users, since cooperation will reduce tolls. The reason is similar to the logic of vertical integration in a supply chain: Non-cooperation would mean that trains and road vehicles have to pay monopoly mark-ups in consecutive countries. Eliminating this double marginalisation would increase overall demand.

The analysis also demonstrates that infrastructure investment and maintenance in one country affect the welfare also in other countries. This implies that not only pricing, but also decisions concerning infrastructure investment and maintenance made on the local/national level, may be affected by cooperation.

But for cooperation to occur, it is not sufficient that the total welfare level increases compared to non-cooperation; the countries must also to be able to agree on how to split the gains from cooperation. If not, they have no reason to establish a club.

The analysis shows that this can not be facilitated by way of the simplest type of agreement – the same toll being levied in all countries and each country keeping the revenue collected. There are different reasons for this for the respective type of network:

• In the parallel case total user costs, including the toll paid, must be equal for every link; if not, users would prefer the alternative with the lowest (generalised) costs. But since other user costs vary, for instance due to that the alternatives are not equally long, so must also the tolls. Uniform tolls would then be impossible.

• In the serial case it is not realistic to expect a country with a very costly link – for instance a road or railway where extensive tunnelling or bridge construction is required – to accept setting the same toll as a country where the link traverses land where construction is easy and cheap. Again, countries would not come to agree on a uniform charging policy.

Some other mechanism than a uniform toll therefore has to be engineered. It is obvious that no correct or efficient allocation exists; once construction of an infrastructure project has commenced, the precise sharing of the costs sunk in the project does not affect the fact that the link has been built. On the other hand, some allocations of costs and revenue are more likely to be accepted by the parties to the deal than others. This is important for the efficiency
issue if it is beneficial for overall welfare to promote deals by designing benefit or cost sharing principles that increase the chances for realising such projects.

Allocation rules that satisfy intuitive properties linked to fairness would therefore increase the chances that mutually beneficial cooperation will occur. The core of the contribution provided by the WP2 work on this issue is to identify such rules for the parallel and serial cases, respectively.

In the parallel case, it is demonstrated that the Shapley value provides a widely acceptable set of rules for sharing the benefits of cooperation. This is one approach to the fair allocation of gains obtained by cooperation among several actors. The setup is as follows: a coalition of actors cooperates, and obtains a certain overall gain from that cooperation. Since some actors may contribute more to the coalition than others, the question arises how to distribute fairly the gains among the actors. Or phrased differently: how important is each actor to the overall operation, and what payoff can they reasonably expect?

The rules which build up the Shapley value are inter alia based on efficiency (here meaning that no resources are wasted), anonymity (two identical players or participants in the coalition are treated equally) and a dummy property (meaning that a player with a constant marginal contribution to every possible coalition of which he or she could be a member, is allocated this constant). The Shapley value represents each player’s bargaining power in terms of a percentage of the total value created. Bargaining power thus varies with value contributed. Persons who contribute more receive a higher percentage of the benefits.

The way in which the Shapley value could be applied to cost sharing in rail infrastructure may be illustrated by the following hypothetical example.

Example: A certain rail line is used by freight (F), regional passenger (R) and high speed passenger (H) services. There is a certain cost associated with providing a railway suitable for freight services only. On top of that, adding a regional service may add some more costs and adding a high speed train would cost even more. Alternatively, the situation could be that the line is first prepared for high speed services. Adding a regional train would probably not cost much extra while the line might require further strengthening of bridges etc in order to allow for also freight services. In total, there are six different sequences of patterns of costs; (FRH), (FHR), (RHF), (RFH), (HRF), (HFR).

Each type of service should clearly pay their marginal costs. But how should the fixed costs be allocated, given that all three categories of trains will be operating? The answer provided by the Shapley value is to calculate the cost for each participant under the respective sequence and divide this by six, i.e. the number of sequences. In this way, the payment of each party which meets the above criteria will be estimated.

There is one proviso, however, namely that the cost share can not be larger than the operators’ respective benefits. Even if costs are shared in the best possible way, as defined by the

---

3 It is possible to make a distinction here between long-term marginal costs, which may be attributable to one particular service, and genuinely common costs which have no such element in them. This distinction would, however, not change the nature of the argument.
The Shapley formula, this does not guarantee that it is worthwhile for one (or indeed any) of the parties to join the club.\textsuperscript{4}

The Shapley value is, however, not appropriate for the serial case. The reason is that it coincides with setting a uniform toll level and letting each country keep their toll incomes. It has already been argued that this would not to be a reasonable allocation. Instead, the analysis considers three alternative allocation rules.

Instead, three new allocation rules are introduced. The proportional rule (PR) ascertains that the profit from cooperation is divided proportionally to the cost each country has for its part. The adjusted proportional rule (APR) means that a part of the profit is divided equally among the players and the rest proportionally to the cost of each country. Finally, and the adjusted equal profit rule (AP) means that each country is first compensated for its costs, thereafter the rest of the profit is split equally among the countries. All three rules allocate more to a country with large costs than to a country with low costs, which is reasonable for the serial case. They also have a number of other nice properties.

It would take the discussion too far to go into the details of these solutions to cost and/or benefit sharing in the parallel and serial cases. It is, however, obvious that the background paper in appendix D demonstrates that it is indeed feasible to provide general ex ante rules for benefit sharing which will boost the chances for a coalition to be formed. This provides further substance to the idea of applying (cooperative) club theory to specific applications for infrastructure provision. We therefore now turn to one possible example of how this type of analysis can be applied, the Eurovignette rules.

5.3 Is Eurovignette a Club?\textsuperscript{5}

In 1993, and after many years of discussion, the Council adopted a first directive on Heavy Goods Vehicle Taxation. One part of the directive was to regulate the minimum level for annual vehicle taxation. A second part was to regulate the level of tolls and a third to introduce user charges and a framework for sharing revenue from these charges between countries. After the introduction of the directive, a number of non-toll countries set up a common system for user charges under the name “Eurovignette”. The Directive is now about to be updated. The purpose in this section is to discuss whether this Eurovignette cooperation could be interpreted as a club, and if so, which observations could be made from the perspective of this theory.

Member States have two choices for a charge based on territoriality (i.e. to charge where the vehicle is driven), and that is making tolls depend on distance or on time driven. The Eurovignette user charge is based on that an annual (or monthly, weekly or daily) lump sum payment is made which entitles a heavy vehicle to use a specific part of the national network, typically a country’s motorways or the TransEuropean road network (TEN). One underlying thought behind this construction is that heavy vehicles require better roads than if a road was built for passenger vehicles only. They are therefore obliged to pay for this extra cost or not be allowed.

\textsuperscript{4}Herein lies a problem with the approach in these circumstances. Any individual operator may be able to reduce their share of the fixed costs by concealing their true willingness to pay. Of course, if all operators pretend zero willingness to pay then no infrastructure will be provided, which may encourage a co-operative outcome. But in practice operators differ in size and in commitment to the route, so in the case of some operators deception may still be worthwhile even if it sometimes leads to non-optimal infrastructure provision.

\textsuperscript{5}This section is based on input from Gunnar Lindberg.
to use these roads. The annual charge is €750 for vehicles with two or three axles and €1 250 for vehicles with four and more axles.

All vehicles using the network, both national and foreign, are obliged to pay the charge. A Member state may make the (annual) user charge an mandatory part of the annual vehicle tax which is regulated with a minimum level. A trucker who pays this fee in one country will not have to pay again for using the dedicated roads in another member country. Moreover, members share the revenues based on the actual vehicle-km driven by vehicles of different categories and nationalities on each other’s networks.

To provide an example of how the system works, assume that the total revenue from the Eurovignette in country \( i = 1, \dots, n \) is \( R_i \). Assume that countries \( i \) and \( j \) are part of the Eurovignette deal. The net payment \( F_i \) to country \( i \) from country \( j \) is \( F_i = k \times (k m_i - k m_j) \). The constant, \( k \), is the annual lump sum charge divided by average distance driven by a vehicle (assumed to be 130 000 km), i.e. \( k = €0,01 \) at the moment for the large vehicles. The number of kilometres driven on country \( i \)’s network by vehicles from country \( j \) is \( k m_i \), and correspondingly for \( k m_j \). The substance of this sharing rule is that country \( i \) will receive money from country \( j \) if country \( j \)’s heavy vehicles on average drive more on its roads than the other way round.

Except for the possibility to get revenue from other members’ use of its own roads, members will also split the revenue from charges paid by vehicles from countries which are not part of the Eurovignette. The way in which this is done is demonstrated in table 1. The split was adjusted when Sweden entered the agreement (in 1997) as well as when Germany introduced its distance based toll system and therefore opted out of the Eurovignette (year 2005).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>73.00</td>
<td>69.16</td>
<td>Not member</td>
</tr>
<tr>
<td>BE</td>
<td>13.00</td>
<td>12.31</td>
<td>39.92</td>
</tr>
<tr>
<td>DK</td>
<td>4.00</td>
<td>3.79</td>
<td>12.29</td>
</tr>
<tr>
<td>LU</td>
<td>1.00</td>
<td>0.97</td>
<td>3.14</td>
</tr>
<tr>
<td>NL</td>
<td>9.00</td>
<td>8.52</td>
<td>27.63</td>
</tr>
<tr>
<td>SE</td>
<td>Not member</td>
<td>5.25</td>
<td>17.02</td>
</tr>
</tbody>
</table>

Source: Sveriges internationella överenskommelser 1997:50

The Eurovignette system may be seen as a club with States as participants. Participation is voluntary and parties will share the gains according to a pre-defined formula.

The logic of membership can be described from the perspective of the two existing club members, \( i \) and \( j \). Vehicles from countries \( k \) and \( l \) are also using their roads, meaning that total revenue of country \( i \) \( (T_i) \) is \( T_i = R_i + F_j + f_i \times (\sum_{j \neq i} k \times km_j) \); revenue from its own vehicles \( (R_i) \); a (positive or negative) transfer from/to country \( j \) \( (F_j) \); and country \( i \)’s share \( (f_j) \) of revenue from countries which are not members (see table 1). The cost of country \( i \)’s hauliers is \( K_i = R_i + G \); where \( G \) is the toll or user charge paid by domestic hauliers driving on non-Eurovignette members’ network.

Admitting a new country \( m \) into the club would mean that the incumbent country \( i \) loses revenue \( f_i \times k \times km_m \). In addition, \( f_i \) would fall to \( f_i' \) since revenue from non-member countries
would now have to be shared also with country m. However, the haulier industry in country i would gain as their cost for user charges or tolls on country m’s network (G) would disappear and they now only pay the user charge once; \( K_i = R_i + G - A \) where \( A \) is the charge previously paid in country m.

<table>
<thead>
<tr>
<th>i and j has formed a Euorovignette cooperation</th>
<th>State m enters the cooperation</th>
<th>Net effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax revenue of country i ( T_i = R_i + F_j + f_i \sum_{k} k \cdot km_i )</td>
<td>( T_i = R_i + F_j + f_i \sum_{k} k \cdot km_i )</td>
<td>( dT_i = -f_i \sum_{k} k \cdot km_i \cdot k' \sum_{k} k \cdot km_i )</td>
</tr>
<tr>
<td>Haulier cost of country i ( K_i = R_i + G )</td>
<td>( K_i = R_i + G - A )</td>
<td>+ A</td>
</tr>
</tbody>
</table>

It is not feasible to say anything about tax revenue for an entrant country, since different countries may use different systems before joining the club. However, it is clear that the hauliers of the entrant will have lower costs. Revenues for incumbent countries will be reduced for two reasons. Country m’s vehicles which previously paid for using country i’s roads will not do so any more and in addition, revenue from non member countries will now have to be split with yet another member country. The up-side of the deal is that country i’s own vehicles will have their costs reduced since they don’t need to pay for using the roads in country m any more.

The only argument for the incumbent country to accept new members is related to the benefit of the haulier industry; taxpayers will undoubtedly loose. The haulier industry will also gain in the country joining the club while it is not possible to say anything about the effect for taxpayers in joining countries. The conclusion is thus that the haulier industry, both within and outside the Eurovignette club, has incentives to lobby for as large club as possible. This conclusion follows well our perception of the process for the development of the Eurovignette club – Member States tries to join the club to please their haulier industry.

5.4 Sharing costs in Swedish road ownership associations

Sweden is a sparsely populated country with large regions with a small number of residents only. While the government is committed to make a road network available for use in all places covered by a city plan – which includes most conurbations of some minimum size – there is no legal requirement to do so outside these areas.

Today’s road network comprises some 98 000 km state roads, 37 000 km local streets handled by local communities (the public sector’s third tier) and 284 000 km private roads. Some 150 000 km of the private roads are built by property owners in order to facilitate the transport of timber from where it has been logged. These roads are typically not maintained between major logging events but are open for the public.

In total 60 000 single individuals, property owners and private road associations (samfällighetsforeningar) own the rest of the private roads. Of these, 24 000 road associations receive state subsidies towards covering parts of the costs for maintenance and construction. State support is conditioned on that the roads are held open for use by the public.
Each road association has only a few members and controls one or a small number of road links. An association is established by a formal procedure under the auspices of Swedesurvey, the national land survey of Sweden (Lantmäteriet). A framework of rules for membership and alternative formulas for sharing costs between members of the organisation are established by law. Within this framework, the road association and Swedesurvey can, however, adjust and design sharing formulas which fit each specific situation.

Property owners may opt out from membership in the association but still be allowed to use the road against a charge related to annual wear and tear caused by expected traffic to and from that property. If the property generates traffic which causes particular damage to the road, the owner is also compelled to pay for renovation. This opt-out opportunity makes the road association into a club according to the definition given above since membership is voluntary and since members have a net benefit from joining: it is cheaper to be a member and to be able to use the road rather than to pay per use or not to get access to the property at all.

In a separate paper for WP2 included as appendix E, the practical problem of how a road association can divide the costs among its members in a fair way has been analysed. The establishment of fair cost-sharing rules would enhance the chances that a road association is established in the first place, i.e. that a club is organised. Clubs organised to handle the dense parts of networks could be seen as a mechanism to organise the provision of rail and road services when a fairly small and reasonably homogenous group of users benefit.

The costs to be shared are of two types. The first category is directly related to using the facility. Actual use is not measured but rules-of-thumb for how different types of properties generate traffic are available. A summer house is for instance assumed to generate a certain number of annual trips and a permanent residence another. It is therefore feasible to let every user pay roughly the costs that they generate. The second category concerns construction and maintenance costs related to weather rather than use. The paper deals with the sharing of this second type of costs making use of the Shapley value.

In the same way as discussed in the previous section, there is no correct or efficient way to allocate costs of this nature. Some allocation principles are, however, more likely to be acceptable than others, which make them attractive for the establishment of a road association. In the railway sector, the association could correspond to the joint construction and maintenance of tracks in industrial areas.

Appendix E demonstrates that a road association is a more general version of what the literature refers to as an airport game. This is a game to share the costs for financing a take-off and landing strip where different parties to the coalition – different airlines using aircraft of different size – demand different lengths of strip.

The road association may be made up of several links where every link of the network corresponds to an airport game. Moreover, the demand for each link by property owners may have different degrees of sophistication. While some members only need access to their cottages during the sunny season, others live there around the year and require winter maintenance. Some estates only use light vehicles while others are accessed by trucks and therefore require a stronger sub-structure. It is obvious that accommodating a user with a certain need allows accommodating all users with a lower need at no extra cost.
The Shapley value of a cost game allocates to each player the marginal cost contribution that the player makes to each of the coalitions to which he belongs. The marginal cost contribution in this context is the cost the player adds to the coalition if he has a higher need for sophistication than the rest of the players of the coalition. It is thus demonstrated that it is feasible to construct cost sharing rules which increases the chances to establish coalitions which voluntarily would make infrastructure available in an efficient way to society at large.

6. Conclusions
A market economy is believed to make use of a society’s scarce resources in an efficient way. Important preconditions for the overall belief in the marvel of markets are, however, violated in the transport sector, requiring the government sector to intervene in one way or another. This is indeed obvious from the extensive involvement of governments in the provision of road, rail, port and airport infrastructure.

There are three motives for government intervention.

- Natural monopolies: If the government would not interfere, there would be too little infrastructure provided at too high costs.
- Externalities: If the government would not interfere, too much traffic would produce too much negative side effects in the form of emissions and noise.
- Equity: In the absence of government intervention, markets may not deliver infrastructure which is equitable. For instance, regions with few inhabitants may have less infrastructure that would be seen to be acceptable from an equity point of view.

Since long, it is accepted that marginal cost pricing is a way to guarantee that the infrastructure that has been built is efficiently used. If each vehicle pays for the costs it gives rise to, including the external costs, users of roads, railways, ports and airports would only do so if their benefit would be at least as large as this cost.

It is, however, also well known that marginal cost pricing in a natural monopoly would not generate enough revenue to cover the costs of providing these services. The reason can be found behind the natural monopoly concept: It costs a lot to have the infrastructure built, but once there, maintenance costs etc. – i.e. marginal costs – are low and would not suffice to recover also original investment costs.

Governments may, however, not accept that revenue from infrastructure use, is smaller than costs for infrastructure provision. It may not be seen to be equitable to have tax payers foot the bill for infrastructure provision, and there are indeed also other motives to recover more than marginal costs. Irrespective of motive, this deliverable addresses this conflict of interest: on the one hand, it is efficient to charge only marginal costs; on the other hand governments may want road users, railway companies, airlines and shippers to contribute towards the full recovery of costs.

A first proposition of the work is that insufficient cost recovery is not of immediate relevance for roads, ports or airports. There are, indeed, difficult tradeoffs to be dealt with also in these

---

6 Four of Catrin’s other work packages are concerned with the empirical estimation of marginal costs for roads, railways, ports and airports.
modes, some of which are discussed in appendix A. It is, however, no reason to believe that these modes are used below capacity due to overcharging.

But the cost recovery issue is a concern for the railway sector. It is demonstrated that some member countries indeed have accepted the Union’s official policy that their railways should not be asked to pay the full costs for building and maintaining railway infrastructure. As a consequence, these governments provide upfront subsidies which have provided for growth in rail travel and transport. Other member countries formally require their national railway monopolies to pay for the industry’s full costs. As a result, the annual reports are in the read, maintenance and upgrading of tracks and rolling stock lags behind and ridership and freight volumes shrink. The policy may also lead to the closure of branch-lines which still have a social value.

A second proposition of this deliverable is that current estimates of marginal costs, and indeed actual practice in many countries, fails to recognise that parts of the railway network, during parts of the day, week or year, are heavily used. One way to enhance cost recovery without jeopardising efficiency would therefore be to implement congestion and scarcity charges.

Charging for scarcity was not necessary as long as the rail industry was being operated as a monopoly. The reason is that one and the same company had to internalise the full consequences of all scheduling decisions and of priorities given at delays. The situation is drastically different in a deregulated railway industry with many different operators. Different techniques are available for internalising congestion and scarcity using prices, but the report’s focus is not on make recommendations for the choice of method. The essential argument is rather to establish that pricing the use of crowded railway lines would increase both efficiency – by providing priority to the most valuable services and by giving incentives to smarter ways to use existing capacity – and equity interpreted as cost efficiency.

If the railway industry is required to increase its degree of cost recovery above marginal costs – including scarcity – the third proposition of this deliverable is to use the acknowledged ways to do so. Over the years, economic theory has developed rules of thumb to raise revenue in a way which minimises any efficiency distortions. Ramsey pricing means that a mark-up could be made on marginal costs in order to generate revenue. The mark-up should then be higher on operators which have poor alternatives to rail, while operators which could readily switch to other modes should be charged a smaller mark-up. A complementary policy is to use a two-part tariff. Except for paying for the marginal cost, the operator would be required to pay a fee for having the right to run the services. This is particularly relevant in a system with competition for the market, while it is trickier to use in a system with competition on the market.

Several versions of these two techniques are available and a long range of experiences can be drawn on if an infrastructure operator would be required to charge more than marginal costs. It is, however, important to emphasise that while these mechanisms would contribute to the minimisation of disturbances, the disturbances would – except for in very special situations – not be eliminated. The manifested consequences would be seen in losses of ridership and freight volumes.

In practice, it is difficult to calculate marginal costs, first and foremost since railway administrations don’t compile information about cost and usage in a way which facilitates the estimations. It is equally difficult to provide precise formulas for how much higher than marginal
costs that charges should be set. It is, however, an obvious recommendation that each country should establish rules for the degree of cost recovery *ex ante*, where after its railway administration could work out the appropriate charging schemes.

Except for this mainstream line of analysis, this deliverable has also considered an altogether different approach to organising infrastructure activities. The idea is thus to reflect on whether infrastructure users should be asked to establish a *club*. Club members would then decide about how much infrastructure to provide and how to charge for it. The research issue is if this decentralised approach to infrastructure provision would overcome the trade-off problem, i.e. if a club would both deliver an efficient total supply of roads, railways, ports and airport and to charge for its use in an efficient way.

A fourth proposition of this deliverable is that this would probably not be a good idea for the road sector as a whole. The reason is the messy decision structure of the “road club” with all road users as potential members. The prime benefit of the archetype club is to make it feasible for members to negotiate over the size and pricing of the network. The difference between meetings of the club’s elected representatives and the convention of a country’s parliament would be small. If a parliament has problems to establish an optimal trade-off, it is not reasonable to expect that it would be easier for a road club.

The corollary to this fourth proposition is that club concept may have some relevance when it comes to the provision of ports, airports and railways. In particular, it is obvious that the number of potential members in these cases is more practical than in the road sector application. The club members as a collective could therefore take votes on the size of the club – the extent of the respective infrastructures – as well as the charges for using it. Non-members could readily be excluded and might even be offered to use the facilities on a limited scale at negotiated prices.

The prime challenge of the club solution lies in its incentives to protect the monopoly position it may have. Even though there might be capacity available to accept entrants, this could increase the competitive pressure, which would benefit society at large but would not sit well with the incumbent club members. Clubs may therefore stifle competition and also in other ways hamper efficiency in infrastructure supply.

There may be ways to reduce the risk that clubs have these negative consequences. The government could, for instance, establish framework rules for when and how a club should be obliged to accept new members. Since there are no free lunches, it should also be acknowledged that some degree of inefficiency may be acceptable if the club has balancing beneficial qualities.

A fifth proposition of this deliverable is that club theory in general, and cooperative game theory in particular, may offer some insights into the organisation of clubs for providing specific packages of infrastructure services. Two possible applications discussed in the report are the Eurovignette directive and Sweden’s private Road Associations. Both applications describe situations which may be interpreted as clubs.

The basic idea is the following: There are situations where it is obvious that two or more countries would both/all benefit from cooperating, i.e. that it would be feasible to improve welfare by cooperating in a club. It is, however, less obvious how the benefits from coopera-
tion should be split between the parties. The inability to agree on how to split the gains may altogether block the deal.

However, by using what is known as cooperative game theory applied to a club, it is feasible to suggest simple rules for sharing the gains. By agreeing on rules and principles that have a general appeal before calculating precise numbers, it may be easier to implement mutually beneficial deals.

Several specific examples of such rules are discussed in the report. It is at the same time only the first attempt to apply club theory on analysing these situations, and more research is required to provide more precise conclusions.

To summarise, the efficient provision and use of transport infrastructure provides huge policy challenges. The present report has dealt with several of these issues but has also had to leave out others. For instance, we have described the need to make the treatment of greenhouse gases in the transport sector part of an official policy, but have not considered the appropriate way to do so and what that might mean for governments’ tax revenue.

The focus has rather been on the risk for insufficient cost recovery if marginal cost pricing is used. The analysis pushes important tasks back to the political level. In particular, it must be a policy decision to establish which financial target should be set for the rail industry in beforehand. Once this target has been established, the report has, however, identified the principles for implementing a policy which minimises the distortions from prices above marginal costs.

The report also opens the door for further analysis of another approach to these issues. Rather than expecting the public sector to be fully responsible for providing infrastructure services, the possibility to establish clubs to do so has been addressed, as well as the use of cooperative game theory to overcome some of the obstacles for club members to cooperate. To be sure, these principles are no panacea, but they may offer insights which were previously not considered.

References
Appendix A. Aspects on marginal social cost pricing and cost recovery in the European transport sector

Magnus Arnek, VTI

1. Introduction

A key ingredient in the European Union’s programme on transport pricing is internalization of external costs. These are costs which are induced by transport users or producers but not borne by them. Instead the external costs are passed on to third parties and the general public. External costs include congestion costs (time losses, vehicle operating costs), accident costs, air pollution costs, noise costs and costs of climate change. A pricing scheme where transport users and producers have to take these external costs into account is likely to both improve economic efficiency and equity. This type of pricing goes under the name marginal social cost pricing. That marginal cost pricing leads to economic efficiency is a well known fact from microeconomic theory. That marginal cost pricing may be beneficial from an equity perspective is given by the polluter-pay-principle, or more generally, that those who use a transport facility and by assumption also benefit from it the most, also are the ones who pay the costs. These two reasons are important motives for why the European Union work towards implementing a transport pricing policy based on marginal social cost pricing.

Implementing this kind of pricing in the European transport sector is a complicated task. A first potential obstacle is the considerable differences in organizational structure, economic conditions, and policy objectives that plague current practice. Regarding organizational structure there are large differences both between transport modes and member countries. For example, some major airports are private while the majority of airports are owned by public sector. The same thing is true for economic conditions: for example there are considerable cost structure and demand differences between and within modes. Finally there are differences in the objective of pricing with regards to factors like financial cost recovery, efficiency etc. All these factors must be taken into account when designing policy and are likely to result in different policy recommendations.

A second potential obstacle is the risk that social marginal cost pricing lead to financial losses and create need for further subsidies and possibly higher taxes. It is a fact that transport facilities like roads, railroads, ports and airports have cost structures that involve a large fixed cost component. This cost structure give rise to considerable economies of scale. In turn this means that marginal cost will be lower than average cost. In other words, charging user fees reflecting marginal costs imply a situation where revenues fall short of costs: the fixed cost component will not be covered. But this traditional view on natural monopoly pricing may not hold true once external costs enter the picture. Revenues from corrective pricing aiming at internalizing the negative externalities can be sufficiently large to cover fixed costs. Hence, no financial deficit may occur. This issue requires further attention.

A third obstacle is that implementation of social marginal cost pricing requires a good understanding of the relevant external costs. In particular, it is important to distinguish between so called pecuniary and non-pecuniary externalities. It is only the latter that should be the objective of corrective pricing. There are some indications that current practice includes external
costs that really are of a pecuniary nature (Baum, et al. (2008)). Furthermore, implementation of social marginal cost pricing requires knowledge of the shape and position of these various marginal costs curves. This is knowledge that does not come for free because these curves are not easily observed and have to be estimated from data that often have poor quality. Hence, estimates may not be very reliable. In addition, it is not always clear what the best estimation method is (Bruzelius (2004)). Finally, externalities are not the only source of market failure in the transport sector: market power and public goods are two other market failures. According to the theory of second best, an efficient pricing policy must also take these two potential efficiency problems into account.

In sum, implementing social marginal cost pricing in the European transport sector is a complex and far from uncomplicated task. How should pricing schemes be designed for various modes? To what extent would social marginal cost pricing lead to lack of cost recovery? Is it problematic in first place, to not have full cost recovery for a certain transport modes, like rail for instance? More generally, what are the main inefficiencies in todays European transport sector? Are negative externalities really the major problem? These are examples of questions that will be addressed in this paper.

The purpose of the paper is to contribute to a broad understanding of the opportunities for and implications of implementing social marginal cost pricing. The analysis will be based on a public finance perspective. By this we mean that it is firmly based in modern microeconomic theory and welfare economics but that we at the same time keep an eye towards current practice.

The rest of the paper is organized as follows. Section 2 presents a theoretical framework Section 3 gives an overview of how the transportation infrastructure is currently organized and financed in Europe. In this section we also discuss potential inefficiency problems in current European infrastructure policy and implications of a cost requirement. In section 4, a few concluding remarks are made.

2. Theoretical framework

This section lays out an analytical framework based on microeconomic theory and welfare economics. We start by briefly describing the economic feature that characterize transport infrastructure. Then we discuss efficiency and equity aspects on infrastructure provision. The discussion centres around three sources of markets failures: increasing returns to scale, externalities and public goods. In addition, we discuss the trade-off between efficiency and equity that nearly always face decision makers in the public sector. The chapter concludes with a discussion about financing principles and the issue of cost recovery.

2.1 What characterise transport infrastructure?

Infrastructure defined

The word infrastructure is a composition of the latin words *infra* and *structure*. It was first used by NATO in the 1950s referring to military base functions like airports, telecommunication-

---

tion systems etc. It has since then been used for referring to facilities like roads, railroads, ports etc.

A transport infrastructure system usually consists of a fixed network of nodes and links upon which a flow of people, goods or information are transported. The characteristic of the flow varies between different transportation modes. Common for all modes, however, is that the higher, the smoother, the more uninterrupted is the flow etc., the more valuable are the services that the system provides.

Some transport systems are more sensitive to disturbances than others. A railway system, for instance, is very sensitive to disturbances. A disturbance in one part of the network spread easily to the rest of the system and prevent it to work properly. The various technical parts in a railway system must fit precisely together if the whole system is going to work well. For this reason railway systems usually require a high degree of centralization. In comparison, a road network does not require the same fit between the parts and is not as vulnerable to disturbances. Therefore the centralization requirement is not equally strong for a road system.

Transport infrastructure systems have different forms. The two most common are:

- The point shape. These systems are accessible for the users form only at the nodes. Examples are airports, ports and railways.

- The line shape. These systems are accessible for the users along the linkages, e.g., roads.

**Transport infrastructure facilities have natural monopoly characteristics**

From an economic point of view transport infrastructure facilities are characterized by the following properties:

- Major indivisibilities.

This means that the system has a minimum scale and typically must be totally completed before it can be used.

- Large sunk costs

A transport system typically involves large investments that to a large extent also are sunk.

- High degree of asset specificity

Many transport infrastructure facilities are constructed in a specific space for a specific use. This means that the resources invested are dedicated to this particular use and have no alternative use. In other words, the resources invested are to a large extent locked into the project.

- Long lived

Infrastructure facilities have usually a very long lifetime. This has several implications. One very important is that the investment must be made with a forward looking perspective.
Together these properties imply the existence of large economies of scale and uncertainties. In fact one may argue that transport infrastructure has the economic characteristics of natural monopolies, a circumstance that make them rather special from an economic point of view, which we return to later.

The government often provides transport infrastructure
In most countries, the government (central, regional, or local) are heavily involved in infrastructure provision. For example, most major roads, railroads, airports, and ports are owned and operated by the public sector. One explanation for this is the technical properties described above.

The public sectors engagement in the transport sector has not always been this extensive. If we go back 100-150 years in time the role of the public sector was much smaller and consequently much more relied on the private sector. Railways are a good example. In the late 19th century most railways around the world were privately owned and operated. Take the USA, for instance. Around year 1900 many private railway companies competed fiercely with each other. At this point in time, the railway industry was large. In fact, some of the world’s largest companies all categories were American railway companies. Over time the situation has changed. The private railways decreased in numbers and more and more were taken over by the government. Today, it is fair to say that the railway sector is dominated by the public sector.

It is important to remember, however, that the government’s involvement in infrastructure provision differ both between countries and between different transport modes.

Well-functioning transport infrastructure is of great importance for economic growth and development of society
The first and probably the most important reason for the government to take on a general responsibility for infrastructure provision is the recognition of how important good infrastructure is for a society’s possibility for development and a belief that the private sector, with its mainly profit oriented objective, is not fully up to the task of delivering this. Fast, efficient and reliable transport-, communication- and energy system, that is various types of infrastructure, has historically had a very large impact on the development of society. It was first when these systems developed that trade and hence economic growth could flourish and develop. In today’s modern highly technological society the importance of functional and reliable infrastructure system is even greater. It is hard to imagine a society without electricity, internet, telecommunications and good transportation systems. The modern society would more or less collapse instantly if these systems would cede to exist. Even minor disturbances could quickly cause loss of production and income.

Governments in democratic countries naturally have broader perspectives on investments in infrastructure than private firms that must earn money on their investments. For this reason government can and will make investments that the private sector would never do.

Finally, the development of large infrastructure systems is associated with large uncertainties that government more easily can handle than private firms because of its size and more diversified business. The uncertainties concern both costs and demand. Construction of large infrastructure systems is plagued with uncertainties about final production costs. It is seldom that the initial budget is kept. In addition, it is difficult in advance to determine the future demand for infrastructure services, not least because the infrastructure is so long-lived.
2.2 Efficiency aspects of infrastructure provision – the risk for markets failures and potential remedies

A main motive for the government’s involvement in the provision of infrastructure is the risk for market failures, which may result both in undersupply of new infrastructure and over- or underconsumption of existing infrastructure and the services produced on it. For most goods and services in the economy private markets can be expected to deliver what society needs and wants and that in a very efficient manner.\(^8\) Infrastructure systems have, however, certain properties that may cause private markets to fail: increasing returns to scale in production, externalities in production or consumption, and public goods features. In this section we look closer into this issue.

**Increasing returns to scale and the natural monopoly problem**

Competition between sellers of goods and services is the key force driving market efficiency. Under certain conditions competition will drive down prices to marginal cost. In markets that are not competitive this pressure is not present and as a consequence an important market failure may occur. In a market where competition is weak or even lacking altogether prices will be too high and consumption too low. The reason is market power – the ability to charge a price over marginal costs. There might be many consumers who are willing to pay a price over the marginal cost of delivering goods to them that are not able to buy the good. This is clearly inefficient.

The degree of market power is, however, not solely related to the number of sellers. It is also related to the demand elasticity faced by each seller in market. The more elastic demand is the less is the seller’s market power and vice versa. This means that even a monopoly may not have market power if the elasticity of demand is large. On the other hand, a monopoly selling a necessity good or service, may have a huge market power.

One reason why markets may not be competitive is the existence of increasing returns to scale in production. This means that average cost decrease with volume, something that calls for large scale production. In this circumstance efficiency requires that there be a limited number of producers. In some industries, called natural monopolies, increasing returns to scale are so significant that only one firm should operate in any region. Transport infrastructure like roads, railroads, airports and harbours are all to some extent examples of natural monopolies.

Roads are a good example. The major cost associated with providing road services, that is, driving on them, is the road network. Once roads have been constructed the additional cost of letting an extra car use the network are small. In fact, the marginal cost is close to zero for passenger cars on a road without congestion. It would clearly be inefficient to have two road networks, side by side, competing to deliver road services to the same potential drivers. The same is true of most railways, ports and airports.

In the presence of increasing returns to scale in production, one can not rely on competition to ensure that the industry operates at the efficient level. As mentioned above, competition tends to drive down price to marginal cost. But if the firm charge a price equal to marginal cost, it will suffer a loss since marginal cost is lower than average cost for industries with declining average costs. Therefore competition is not viable.

\(^8\) The fundamental theorem of welfare economics state that competitive markets produce efficient outcomes.
Policies towards natural monopolies

Many infrastructure facilities have natural monopoly characteristics. Hence competition is not feasible or desirable. Total costs will be minimized with only one supplier. On the other hand, this supplier will have some monopoly power. Economic theory points to several ways to regulate the supplier to curb the problem of market power.

Public ownership

One common way for the public sector to deal with the natural monopoly problem is to own the asset. This way the government has control over pricing decisions, production decisions, locations decisions etc. The most common pricing strategy is to charge prices that cover the public entities own costs. In this way no tax subsidies are necessary. Another pricing policy is to charge marginal cost fees and use tax money to make up the differences between revenues and total costs.

Regulation

A second approach is to leave the natural monopolies in the private sector but to regulate them in various dimensions. Special attention is typically given to the prices that the firms are allowed to charge. The most basic form of regulation is marginal cost pricing. The government insists that the monopoly supplier charges a price equal to marginal cost. If the production gives rise to external effects the price should equal the marginal social cost. A subsidy equal to the difference between revenues and total cost will make up the loss.

However, marginal cost pricing regulation and the need for subsidies ignores the question of how the revenues required to pay the subsidy are to be raised. This solution also presumes that the government knows the size of the subsidy that will enable the infrastructure provider to be viable. An alternative to marginal cost pricing regulation is average cost pricing. Here the government insists that the infrastructure provider charge a price equal to average cost. This way, the infrastructure provider pay for itself and no subsidy is necessary. Average cost pricing reflects a compromise between a desire for efficiency – setting price close to marginal cost – and the need for the infrastructure provider to break even. Marginal cost pricing typically means that general public pay to subsidize a private good that is enjoyed by only a portion of the population, or enjoyed by different individuals to different extent. This is not so under average cost pricing regulation.

More advanced pricing schemes

So far only unit price regulation has been considered: the price per unit of the service consumed is the same – equal to marginal cost or equal to average cost. A more general pricing scheme is two-part tariffs, or linear pricing. This means that the price the consumer faces for a unit of the product varies with the quantity chosen. With a two-part tariff the consumer is required to pay a fixed charge in order to consume the service at all and in addition a marginal price per unit of service consumed. In general, a two-part tariff is welfare improving. By charging a price per unit of some service consumed and a fixed fee to cover fixed costs, efficiency is achieved at the same time as revenue is generated to cover cost.

Another pricing scheme of interest for this project is Ramsey-pricing. This is a pricing technique normally associated with multiproduct natural monopolies. Ramsey pricing is the answer to the problem of maximizing efficiency under the restriction that revenues must cover cost. Suppose that consumers demand for the supplier’s products are independent. In this case, products for which demand is relatively high should have higher mark-ups over mar-
original cost than products for which demand is low. Ramsey pricing does not require that common cost should be allocated to individual products.

One special case of Ramsey-pricing, called Peak-load pricing, can be used when demand fluctuates systematically over time. For example, people usually travel to work by road or by rail tend to do so between 7-9 and 16-18, i.e., during rush hours. The major parts of roads or railways capital costs are incurred in construction phase. During peaks, when demand is high, the marginal costs of providing peak services are much higher than the marginal cost of the off-peak service. The reason is, of course, that the latter service is possible to supply using the existing capacity, while this is not so for the former. Therefore the Ramsey-formula for the optimal pricing suggests that prices for utilization of the infrastructure should differ between peaks and non peaks. The price during the peaks should be higher than the non-peaks with the purpose of reducing demand for costly peak services.

Promoting competition
The third strategy to deal with market power is to encourage competition. This is not an easy task given the natural monopoly characteristics that many infrastructure facilities have. However, technological advances and new management ideas may change the perception of what a natural monopoly is. For example, railways have for long time been a textbook example of a natural monopoly. But upon further examination most of the scale economies belongs to the infrastructure, not the actual operation of trains. This insight has resulted in an unbundling of infrastructure from operation. Competition is promoted in latter activity. Another example of introducing competition is competition for the market instead in the market. This is handled by way of competitive bidding.

Externalities

Externalities defined
A second type of potential market failure associated with transport infrastructure is externalities. Externalities arise whenever an individual, firm, government agency, or country carry out an action that has an effect on another individual, firm, agency or country, for which the latter does not pay or is not paid. Externalities can in principle be both positive and negative.

The problem with externalities is that they create a wedge between the private benefit or cost of doing something and the social benefit and or cost. When this happen individuals taking decisions to maximizing utility, profit etc, tend to make inefficient decisions. In case of an activity giving rise to a negative externality, there will be overproduction or overconsumption of that activity, while there will be undersupply or underconsumption of goods or services generating positive externalities.

The distinction between pecuniary and non-pecuniary externalities
In a modern society where interactions between individuals and firms abound, there are many situations where decisions taken by one individual would affect another individual’s well-being in a way that the former does not take into account. Competition between firms in a market is one example – a restaurant establishing in a new area may not take into account that it hurt business by already established firms. But this type of externalities is not a source of market failures. The reason is that the external effects are reflected in changes in market prices. Externalities that work trough the price systems are called pecuniary externalities. It is the other type of externalities – non-pecuniary – externalities that is a source of market failures – that is cause inefficiencies in the form of under- or overconsumption. It is sometimes
hard to determine whether an externality is pecuniary or not. In Chapter 3 we will see that some external effects that the European Commission seems to regards as non-pecuniary actually might be classified as pecuniary. This issue is clearly important given that it is only non-pecuniary externalities that cause inefficiencies that need corrective measures.

*Externalities in the transport sector*

Transport infrastructure potentially creates both positive and negative externalities. New infrastructure typically improves production and consumption possibilities for a large number of individuals and firms. It opens up new possibilities and combinations whose potential may not be seen at the point in time when the investment decision is taken. These effects may be hard to fully take into account, not least because it may be impossible for the investor to capture the benefits by way of charges from the individuals and firms who gains from the investment. Uses of transport infrastructure sometimes give rise to negative externalities in the form of congestion, pollution, noise and accidents.

**Methods of dealing with externalities**

There exists several ways to deal with the problem with external effects. One may distinguish between market based methods and non-market based methods. Examples of the former category are taxes, subsidies and marketable permits. Regulation is an example of the latter.

*Taxes, fees, fines and subsidies*

One way for the government to prevent overproduction of activities that bring negative externalities is to provide incentives for consumers and producers to decrease the activity. This can be done by introducing taxes or fees. Or the government may introduce fines. When the production of good or service creates positive externalities, the government can try to increase the supply by subsidizing the production.

*Marketable permits*

Marketable permits limit the amount of emissions that any producer can emit. Each producer is granted a permit to emit so many units of a certain pollutant. The producers are allowed to sell their permits. A producer that decreases the amount of pollutants it let into the air could sell some of its permits to another company that wants to expand production.

*Regulation*

The government may also deal with externalities by way of regulation. It may set emission standards for example.

*Public goods*

A third type potential market failure present in the transport sector is the public goods problem. Most goods and services produced and sold in the economy are so called private goods. Private goods are characterized by rivalry in consumption and exclusion in consumption. The first property means that one individual’s consumption of a good precludes another individual’s consumption of the same good: in others words, the marginal cost of providing the good to another consumer is positive. The second property, exclusion, means that it is possible to prevent someone from consuming a good. Exclusion is normal in ordinary markets and works through the price system: consumers who are willing and able to pay the going market price obtain the goods, while consumers who are not willing and able to pay do not.
Goods and services for which there are no rivalry and nonexclusive are called public goods. Classic examples of public goods are defence and lighthouses. A distinction can be made between pure and non-pure public goods. Pure public goods are goods for which both the properties above hold. Pure public goods are rare. It is much more common that a good possess either non-rivalry or non-exclusion in consumption.

There are two basic forms of market failures with public goods: underconsumption and undersupply. In other words, for goods which have public goods characteristics, the private sector can be expected to deliver incorrect quantities both on the production and the consumption side. The basic dilemma is this: when a good is non-rival in consumption exclusion is undesirable because it will result in under consumption. Thus, when a good is non-rival in consumption it should not cost anything to consume it – it should be free of charge. But without exclusion, that is, charging a non-zero price, there is the problem of undersupply.

There is no easy way out of this dilemma. The answer is often that the government step in and provides public goods and finance it with general taxes.

### 2.3 Equity issues

The third reason why governments are involved in infrastructure provision is to pursue equity aspects. The private sector may be very efficient but fail to satisfy social standards of equity. Here we briefly discuss equity issues in the transport sector.

**Horizontal versus vertical equity**

Fairness is concept with many dimensions. In the economic literature the following two stands out: horizontal equity and vertical equity. A government program is considered horizontally equitable if individuals who are the same in all relevant respects are treated the same. The principle of vertical equity says that some individuals are in a position to pay more and that these individuals should do so.

**Equity issues in the transport sector**

There exist several equity issues in the transportation sector. For example, in many countries one important political objective is that standards of living and opportunities for individuals and firms should be the same in all parts of the country – a kind of horizontal equity. However, economic conditions often differ greatly between different regions. Some regions are rich with a prosperous business environment and high incomes. Other regions are poor with high unemployment, low economic growth and consequently low incomes. Because of these differences demand for infrastructure services vary a lot. In regions where demand for infrastructure is large also willingness to pay is large. This makes it possible to finance, build and operate new infrastructure like roads, ports and airports with user fees. For example, this is how most large airports are financed. Airline ticket tax, revenues from commercial activities like shopping malls, bars, parking facilities etc. generate revenues that more than cover costs. Thus, government grants are not necessary.

In regions where willingness to pay is low, the opportunities to finance transport infrastructure with user fees are much bleaker. For exactly this reason, the governments in many countries have cross-subsidiation schemes for the national airport systems. Large airports, with large demand and high willingness to pay make profits that are used to finance smaller airports in regions with lower demand traffic. Although, the example here was air traffic, the
reasoning is valid for other modes as well. This is one example where vertical equity aspects come into play.

Another equity dimension concerns user fees. Governments sometimes charge fees to those individuals and firms who benefit from a publicly provided good or service. Examples are toll-roads, ports and airports. User fees have an equity aspect because those who use a public facility also are those who pay the most. For this reason, user fees are an equitable way of raising revenue to finance public facilities.

The efficiency-equity trade-off

Most economists recognize that there exists a trade-off between economic efficiency and equity: more efficiency usually means less equity and vice versa. The reason for this trade-off is that pursuit of equity usually involve some kind of interference in the workings of markets. For example, pursuit of equity usually involves redistribution of resources from those who have to those who have less. This is typically achieved by taxation. But taxation creates distortions in the market.

One example of an efficiency-equity trade-off in the transport sector is related to the distribution system between airports mentioned above. Efficiency requires that airports charge, airlines and other customers prices that are in line with marginal costs. Under marginal cost pricing airports would make only “normal profits”. (In the case of increasing returns to scale marginal cost pricing would not generate enough revenues to cover total costs.) This implies that a necessary condition for a large airport to help smaller airports financially is that the former charge prices or fees that exceed marginal costs. But this practice would create a loss of efficiency.

The practice of charging user fees to finance goods or services with public goods characteristics is another example where pursuit of equity give rise to an economic inefficiency. Some people would say that it is only fair to finance a transport facility with user fees because then those who use the facility are also the ones who pay for it. But for a facility that is a public good, like an uncongested road, a user fee would introduce an inefficiency because the marginal cost of an extra user is zero.

2.4 Aspects on public finance

One of the main concerns of implementing marginal social cost pricing in the European transport sector is the implications for cost recovery. In this section we discuss in overall terms different micro aspects on cost recovery: implications for efficiency, equity etc of a requirement for each transport mode to bear its own costs. We start, however, by a brief overview of the government’s revenues and expenditures.

Public sector revenues and expenditures

In general, governments raise revenues to cover their expenditures in two ways: taxes and fees. In most countries taxes (income taxes, payroll taxes, corporate taxes, excise taxes, etc) constitute the main source of income for the public sector. Among these, the individual income tax and value added taxes typically generate the largest revenues. User fees also generate considerable revenue for the public sector. User fees are used for things as toll roads, harbour, some park and recreational facilities. Other services provided by the public sector fi-
nanced by user fees are garbage collection, water systems, etc. Co-payments for healthcare, childcare, and elderly care are other forms of user fees.

The distinction between taxes and user fees
How can we distinguish a user fee from a tax? A user fee is typically imposed on individuals who use certain services provided by the public sector. In other words, a user fee is a voluntary expense, one that the individual can avoid by not consuming the service. Moreover, a user fee is typically proportional to the use of the service: the more an individual consumes the more she or he has to pay, and vice versa. Finally, a user fee is typically designed to defray the production costs of the public service. A tax, on the other hand, is in general not voluntary. Instead, it is a mandatory assessment on an individual, firm or other entity based on a certain characteristic, such as income, profit, sales, consumption etc. Furthermore, the main purpose of a tax is to generate revenues for the public sector, not defray costs. Last, but not least, a tax is a payment that not necessary give anything in return.

Large variations in how public sector expenditures are financed
There are considerable differences in how various services provided by the public sector are financed. Some services are solely tax financed, like defence, police and basic education. In some countries, like England, healthcare is more or less completely financed by taxes. Other services provided by the public sector are not tax financed at all. For instance, garbage collection is usually financed only from user fees. A third category of services is financed by a mix of taxes and user fees - health care belongs to this category in most European countries.

Behind these differences lay a number of different motives related to efficiency, equity, tradition and culture. For example, provision of public goods is usually tax financed because it is efficient – private provision would result in an undersupply, if provided at all. On the other hand, basic education is tax financed mainly in a pursuit of equity and because of positive externalities from education. The use of user fees can to some extent be explained as an attempt to curb moral hazard and an overuse of publicly provided private goods, like healthcare. The other side of the coin is that user fees may cause equity problem because they may prevent the poor from consuming some good or service.

Government expenditures has risen dramatically over time in most countries
The last 60-70 years, governments expenditures have risen dramatically in almost all advanced countries. Around 1940, the public expenditures as a share of GDP were around 10-20 percent in Europe. Today the corresponding figure is between 40 and 50 percent, with the Nordic countries as the largest spenders. The main expenditure programs are systems for transfers payments, education, health- and elderly care, and interest payments. Considerable change in the composition of expenditures has also occurred during this period, in a direction towards more social spending. This rise in social spending can be explained by three major forces: the democratization movement that started in 19th century, economic growth and longer lives.

Government spending is determined in a political process that by and large reflect the preferences of the majority
As we will return to the merits of earmarking of taxes and calls for balanced budgets between transport modes it is worth remembering that both the level of government expenditures and their composition are determined in a democratic and political process. In other words, how

---

9 This no longer holds in case a two-part tariff is used.
much the public sector spends on various things and sectors in the economy is the result of rather complex political processes where different interest groups promote their questions, fight for funds, where elected politicians after carefully weighing these different interest against each other finally decide on a allocation of funds and what finance method to use. In well functioning democracies, which most European countries are, the outcome of the political budget process, by and large, is likely to reflect the preferences of the people, or at least not deviate to much from it.

*Over time the public sector must run a balanced budget*

No matter how the public sector finances its expenditures – through taxes or fees – in the long run total revenues must equal or exceed total expenditures. For a shorter period of time it may run a budget deficit and finance the revenue shortfall by taking up loans. However, this means higher taxes/fees or lower expenditures in the future. Thus, the government as a whole must run a balanced budget over time.

This is not so for individual activities that the public sector carry out. More precisely there is nothing saying that a specific government activity must cover its costs. This is not strange once one remember that many services that the public sector carries out are not revenue generating – basic education for instance. The expenditures for this category of services are paid by taxes – revenues from taxation of completely different things. A requirement that all services should bear their own costs is therefore unrealistic.

*A current trend towards more financing by fees*

In many European countries there seems to be a trend towards more use of user charges. This is probably the result of the following dilemma. On the one hand, there is a demand for better public services, like healthcare but also transport infrastructure. This can be achieved by a combination of higher cost efficiency, higher productivity and more resources. The last 15-20 years, the public sector has worked intensively along the two first routes, with quite some success too. But these efforts have a boundary, which beyond further improvement is hard. So more resources are probably necessary if demand is to be met. On the other hand, the scope for generating more resources by raising taxes is bleak, especially for the Nordic countries which have the highest tax pressure in the world: around 50 percent of GDP. Further tax raises is likely to have negative effect on economic performance and would not be approved by the majority of the people. This means that the remaining source of revenue is user fees. Hence, the trend.

*Desirable characteristics of a good tax system*

In modern economies few things avoid taxation and consequently there exist a large number of different taxes that generate revenue for the public sector. Tax systems can be designed in a large number of ways. One tax system may be better than another tax system. What is meant by “better” to some extent depends on who you are asking but there exist no unambiguously understood definition of good tax system. However, most economists agree that the five properties listed below are important elements to consider when designing taxes.

The first property is fairness or equity. As already mentioned two commonly used concept of fairness are horizontal and vertical equity. Horizontal equity means that two individuals with the same characteristics, consumption pattern etc. should pay the same amount in tax. Vertical equity means that individuals or firms that are better off, have higher income, profits etc. should pay higher taxes than individuals who are worse off.
The second property is economic efficiency – a good tax system should promote efficiency. Nearly all taxes change individual behaviour and therefore the allocation of resources. Taxes affect incentives to work, save, invest, consumption, etc. In a situation without market failures, taxes in general introduce some form of inefficiency – an excess burden. The art of taxation is to raise a certain amount of revenue with the least excess burden possible. A useful guide is the Ramsey-rule. It says that to minimize total excess burden, tax rates should be set so that the percentage reduction in the quantity demanded of each good or service is the same. The Ramsey-rule can be reformulated in the following way: tax rates should be set inversely proportional to price elasticity of demand. This implies that if the demand for a good is sensitive to price changes, this good should has a lower tax rate than a good whose demand is price inelastic.

In a situation with markets failures taxation may not only generate revenue, it may also improve efficiency. This is the case with negative externalities. Here a free and unregulated market typically produce too much of the activity generating the externality. A tax that increases costs and reduces production therefore improve efficiency.

The third desirable property of a good tax system is that it should be simple to administer. Running a tax system involve both direct and indirect costs. The direct costs are the costs for the tax authority. The indirect cost is the opportunity costs of all individuals preparing and processing tax forms.

The last two properties are flexibility and transparency. A good tax system is easy to adapt to changing conditions. It is desirable that it is visible to tax payers what taxes they are paying and how the system works.

Optimal user fees
Many of the considerations that are valid in the design of good tax system are valid when the public sector should decide on optimal user fees. Economic efficiency requires that the prices users pay to consume a certain service provided by the government – the user fee – is set equal to marginal costs. This way efficiency in consumption is promoted. User fees should also be fair, reflecting both horizontal and vertical equity.

2.5 Aspects on cost recovery
So far we have only touched upon the cost recovery issue by briefly discussing revenues and expenditures for the public sector. In this section, we focus more directly on the issue. The discussion is continually held on a principal level. Practical discussions are deferred to the next chapter.

Cost concepts
Let us begin an overview of different costs concepts. Suppose a government agency is about to build a new road that it then should operate. By definition, the agency’s cost is the total expenses of building the road, operated and maintain it. The expenses come from various sources. In the building process, the agency use a number of resources – factors of production like labour, materials, and capital goods. Labour costs are what the agency pay for its staff. If the agency contracts out some or the work to external firms these costs may also be seen as labour costs. The costs of material include raw material and intermediate goods. The latter consists of whatever supplies the agency buy from external sources. The cost of materials includes the cost of machinery, buildings etc.
The agency’s total cost of building a road are the sum of the costs of the factors of production that are employed in the various stages of production from planning, design to the actual build. To the construction cost we must add operation and maintenance costs. These costs are the expenses on the inputs used in the two processes. Due to the fact that a road usually has a long life span operation and maintenance costs can be substantial.

The total costs are made up of fixed and variable costs. Some of the agency’s costs do not vary with the amount of traffic on the road. For example most of the building costs are fixed because they are incurred before the road is opened up for traffic. Some of the operation and maintenance costs are also fixed. But other inputs are costs that vary with the level of traffic, they are variable costs. The line between fixed and variable costs is not always clear. Maintenance costs for instance have both fixed and variable components. Other costs are semi-fixed: fixed over certain ranges of output but variable over other ranges.

It is important for economic decision making to separate marginal cost from average cost. By definition, average cost is equal to total cost divided by the production volume. The average operation and maintenance cost for road is the total expenses for these activities divided by the traffic volume. Because a large portion of these costs are fixed, the average cost decline when the traffic volume increase. The marginal cost is equal to the change in total costs when output increase by one unit. In the road example, it is the change in total operation and maintenance costs of increasing the traffic volume with one vehicle. In practice, this cost is close to zero, at least for passenger cars.

For a commercial firm marginal and average cost is relevant for different types of decisions. Marginal cost is the relevant cost concept when the firm should decide how much to produce. In order to maximize profit the firm should expand production to the level where marginal revenue is equal to marginal cost. Average cost is the appropriate cost concept when the firm should decide whether to produce at all. In order to make a profit the firm’s average revenue must exceed the average cost. Otherwise, the firm’s total revenues are smaller than its costs, and it make a loss. A firm can not make losses in the long run because it will not have enough money to pay its bills.

A sunk cost is an investment in an asset with no alternative use, which means that is an asset with no opportunity cost. The concept of sunk costs should not be confused with fixed cost. For example, a railroad between A and B needs a locomotive and a crew whether it hauls 500 passengers or 100. The cost of the locomotive is a fixed cost. But it need not be sunk. If the railroad stops trafficking this route it may be able to sell the locomotive to another railway company.

**Full cost recovery is necessary for commercial firms**

A profit-maximizing firm would choose a price to maximize the difference between sales revenue and costs. This means a price equal to marginal costs. A firm that over time continuously fails to take out prices that cover its expenditures will sooner or later go bankrupt because it will fail to pay all of its bills. For a while, the firm’s owner may put in money into the firm if they believe that the profitability opportunities will improve in the future. In other words, full cost recovery is a necessary condition for a commercial firm.

**In competitive markets full cost recovery will be achieved but not more**
Firms acting in a competitive environment will typically be able to cover is costs, including necessary investment costs. But they will typically not be able to earn any rents – that is, sales revenues will be just big enough to cover its expenses. Moreover, this will take a lot of hard work. The competitive surrounding will force firms to exert effort to boost sales at the same time that will seek ways way to reduce its costs and increase productivity.

*The value created in competitive markets typically by far exceed revenues and costs*

The benefits created by markets providing goods and services demanded by customers are often much greater than the revenues firms are able to earn. And consequently, benefits are much higher than the firms total expenditures. This is a fact easy to forget when one discusses cost recovery. Goods and services produced and sold in markets creates benefits or values that not fully are captured by sales revenues. More importantly, from an efficiency perspective it is the size of this value – the difference between the consumers aggregate valuation and the total outlays that is of interest, not to what extent costs are covered. The greater this difference is the more value or surplus is created to the benefits of the individuals in society. This is of course also true for goods or services provided by the government.

*A monopoly firms often more than cover its costs...*

When competitive pressure is lacking, as in the case of a monopoly, sales revenues usually exceed total expenditures. In other words, more than full cost recovery is obtained. In addition, this economic profit or rent may be earned without the hard work that renders competitive firms only zero economic profit. The difference is market power.

*...at the expense of economic value*

The rents that monopoly firms in general earns, with less effort than competitive firms, have a price – less economic value is created. This is one of the major drawbacks of profit-maximizing monopolies. In order to maximize the difference between revenues and costs, a monopoly chooses a higher price than what would result under competition, with the result that less is produced and sold. One implication of this is that there is not necessarily a positive connection between the degree of cost recovery and the economic value that is created. In other words, cost recovery is by itself not necessarily a good thing.

*Price discrimination is in general both value enhancing and improves cost recovery*

From an efficiency perspective, the problem with monopoly is not only due to the lack of competition, it also hinges on the pricing strategy. A monopolist that is able to charge different price from different consumers or different prices depending on the amount of a good or service consumed can often increase revenue compared to using a constant unit price. In addition, efficiency is improved at the same time because the incentives for the monopolist to sell more increase.

There exist several price discrimination techniques. One is to charge different consumer categories different prices depending on some easily observable characteristic like age, location, time of day, week etc. Another is to charge different unit prices depending on quantity consumed. A consumer who buys a lot may receive a quantity discount in the form of a lower price per unit consumed.

From a cost recovery perspective price discrimination is interesting because it may be an effective tool to increase revenues and thereby cost recovery. Moreover, this is not at the expense of the value created, quite the opposite.
Cost recovery in the public sector

Let us now turn to the issue of cost recovery in the public sector. A first question that begs an answer is how we should define cost recovery. For the time being we exclude external costs from the discussion.

A first recognition is that in a free society where all resources are voluntary supplied all costs must somehow be covered – may it be in the private sector or in the public sector. Very few resources that go into some kind of production are free of charge. Labour must be paid, capital must be paid, various resources like electricity, water etc. must also be paid. This is true irrespective of whether a service is financed by taxes or by user fees. The statement that all costs must be paid is almost a truism but never the less is worth mentioning: a production activity in the public sector draws resources precisely as in the private sector. The difference in this respect between the two sectors is how the resources are paid. In the private sector the firm’s revenues come mainly from sales revenue. In the public sector, the finance comes mainly from tax revenues and user fees. Thus, in one sense, the issue of cost recovery is irrelevant – all costs must be paid. This issue is instead how a public service should be financed – by taxes or user fees.

But this is not normally what people mean when they argue for full cost recovery. Instead they mean that cost recovery is defined by the extent to which the costs of publicly provided service are covered by revenues paid by users of the service. The payment may be in form of taxes or in the form of user fees. According to this definition full cost recovery is obtained when user payments completely covers the production unit’s expenditures. At the other extreme, zero cost recovery is obtained when none of the production unit’s expenditures are covered by user fees.

Arguments for balanced budgets...

It is sometimes argued that an implementation of marginal social cost pricing would lead to unbalanced budgets between transport modes. Some people fear that marginal social cost pricing would lead to sharp tax rises on some transport modes, like car use, and that the revenues raised would be used to subsidize deficits in other public transport sectors, especially rail and commuter traffic. For this reason, these people propose that all modes balance their budgets, that is, full cost recovery should be the norm for all transport modes. This proposal for balanced budgets is motivated in the following way.

First there is the efficiency concern that charges below average costs would lead to excessive use of the transport facility in question. Second, there the cost-efficiency concern that the absence of a balanced budgets requirement for a transport mode like rail, which is heavily subsidized, would decrease the incentives for the actors in that sectors to hold down costs, make rational investments, and increase productivity. Third, there is a fear of loss of transparency. Finally, there is a concern regarding unacceptable transfers between population groups, or more precisely a fear that car users would subsidize rail users.

...are weak

What are the merits of these arguments? Note first that there are both equity and efficiency related arguments. Regarding the first argument that marginal cost pricing would lead to excessive use of the transport facilities, it is simply wrong. As already mentioned several times, marginal cost pricing in general is the pricing scheme that leads to efficient use of resources. When users consume a resource up to the point where the marginal benefit (price) is equal to marginal cost, an outcome is reached from which it is not possible to improve the situation for
any of the individuals involved without at the same time make it worse for someone else. Average cost pricing, on the other hand, imply an efficiency loss since marginal benefit usually exceed marginal cost, with the consequence that too little is consumed. Hence, average cost pricing is likely to result in an outcome where transport facilities would be use to little.

The second argument is a perhaps not wrong but not obviously true either. It is true that cost-inefficiency has been low in parts of the public sector, and the railway sector indeed belongs to this group. It is also true that a requirement that any expense must be covered by earnings generated by services supplied, put a limit on what expenditures you can afford. In addition, it may work as an incentive to deliver what the consumers want.

But the argument is still flawed. First, there is strong indication that the cost-inefficiency that has been observed in parts of the public sector is due to lack of competitive pressure, not the source of financing. A balanced budget requirement does not affect that. Second, a system with user fees may actually create inefficiencies under certain circumstances. The monopoly position is likely to lead to too high user fees and in combination with a relatively low price-elasticity the agency is likely to have more revenues than before and may therefore think that they can afford higher expenditures than with a fixed annual budget.

The transparency argument in favour of cost recovery is valid. It would be clear how much each mode really costs and how it is financed.

The final equity argument is weak. As we will see in the next chapter some countries in Europe explicitly use revenues generated in the road sector to finance building of high speed rail lines. Thus, these countries deliberately let car users subsidize rail users. To say that this practice involve an unacceptable transfer from car user to rail users miss the point that these transfers are the outcome of a democratic political process, where elected politicians representing the people, weigh different objectives, interest and aspects against each other. Rail users may argue that it is unacceptable that car users contribute to global warming the way they do.

In sum, the arguments for balanced budgets are weak.

Arguments for and against ear-marking of taxes
Another motive for ear-marking is to secure a budget for necessary investments and to increase willingness to pay higher road charges: people may be more willing to pay road charges when they know that the revenues go to improve the roads. This alleged benefit must be weighted against the potential drawbacks that ear-marking has. One con is that ear-marking may lower financial discipline. A more substantial drawback of ear-marking of revenues is that it interfere in the workings of political process.

What about external costs and cost recovery?
So far we have left out the important issue of external costs and its connection to cost recovery from the discussion. Here we take up this issue.

The problem with externalities is that they, if left uncorrected, will lead to inefficient use of resources. Let us for simplicity focus on negative externalities. Road transport gives rise to a number of external externalities, like emissions of greenhouse gases. The greenhouse gases contribute to global warming. A warmer climate is likely to have a number of negative effects, some potentially catastrophic, on individuals around the globe. If these negative effects
are assigned monetary values it is in principle possible to calculate the external costs that road users are inflicting on other individuals, costs that the former group do not pay and hence do not take into account when they make their decisions on how much to drive, what vehicle to use, etc. As a result they drive too much – more than they would do if they needed to pay the external costs – and consequently let too much green-house gases into the atmosphere.

Several things are worth mentioning here. First, as the example indicates, the external costs are in general a function of the amount of some activity carried out, here driving: they are zero when the cars stand still and increase with the amount of driving. The exact shape of the relationship may vary depending on a large number of factors. Second, the “efficient” amount of driving, and consequently, the efficient amount of externalities, is not zero but occur at the quantity where the road users marginal benefit from driving equal the marginal social cost of doing so (the private marginal costs plus the external marginal costs. Precisely at this point, the magnitude of the external costs that are efficient is determined. It is this external cost that a road user inflicts that is “OK”, or correct, from an efficiency perspective. Note also that the efficient quantities are likely to vary between different users. If we sum all these quantities we get the total, efficient external cost.

When discussing, who should bear these external costs, it is important to not lose sight of the ultimate objective from an efficiency perspective: to reduce the amount of externalities from an excessive level to the correct level. A negative externality can in principle be reduced in two different ways: by a reduction of the activity that give rise to the externality – in the example above, the amount of driving – or by using a technology that allow the same volume of the activity to be carried out with a lower amount of emissions let out. As pointed out in section 2.2, there exist several methods to do this: Pigouvian taxes, regulation and marketable permit systems. Each of these methods can be used to bring down the externality to the efficient level. Here we are interested to what extent these different methods affect cost recovery.

We begin with the case without “internalisation”. Under this regime, the amount of driving and emissions are inefficiently high. The whole external cost is brought on to others, whose welfare is reduced. In this case, the social cost exceed the private costs by exactly the external cost that stem from inefficiently large amount for driving. The car sector does not bear its costs.

Consider next marginal social cost pricing, or the Pigouvian method. Here car users must pay a tax per unit driven, say per kilometer, amounting to the external marginal cost at the efficient quantity of driving. Road users must now take this external cost into account when deciding how much to drive. They find it utility maximizing to drive exactly the efficient level. They “produce” an external cost onto others that is equal to the amount of tax they pay. In other words, the car sector would under the Pigouvian system bear the full social cost that they inflict on society. Note, if the tax is set too low or to high, the amount of driving and pollution would be inefficiently high or low, and consequently the car sector would not respectively would pay too much of the costs.

With a regulation scheme, the government sets a ceiling for how much each car user is allowed to emit. Under this system each car user is forced either to reduce the amount of driving or to buy another car which emits less for the same amount of driving. This method may also reduce the externality to the efficient level. An external cost is still levied on the rest of society, but it is the “correct” one. Regulation does not involve any direct payment from the car users, as with the Pigouvian solution. Thus, in one sense, the car sector does not bear the
full social cost – cost recovery is not achieved. For sure, car users pay a price: they can no longer drive as much as they want or incur expenses for a more fuel-efficient vehicle allowing them to keep the amount of driving up.

 Marketable permits means here that each car user is given a permit to emit a certain level of pollution. The aggregate level is set by the government to the efficient level. Since the government wants to achieve a lower level of pollutants, car user with less fuel-efficient cars must either reduce the amount of driving or buy permits from car users who have more fuel-efficient cars or who do not want to drive so much. A marketable permit system may also reduce the amount of driving and attain the efficient level of pollutants. Again an “efficient” external cost is brought on to society. Car user with fuel efficient cars, that do not need to use their permits do not pay anything, instead they may get paid from selling their permits to car users who need them. In other words, with marketable permits, car users with less fuel-efficient cars pay a cost to other drivers. It is a transaction scheme that only takes place within the car sector. It is unclear, whether the money changing hands are equal to the external costs.

In sum, several methods exists that may be used to deal with a negative externality. Under all schemes an “efficient” external costs are brought onto society. These schemes differ, however, when it comes to the extent that the external costs are borne by the polluters, here the car users. Only with the Pigouvian system, the car users directly pay the external cost – full cost recovery is achieved. With the regulation scheme, no direct payments are carried out. With the marketable permit system, payments are only carried out within the car society.

Several tentative conclusions emerge out of this discussion. The first conclusion is that there seems to no obvious connection between the possibility of achieving efficiency by handling negative externalities and the degree to which full cost coverage is achieved. It is obviously not necessary to achieve cost recovery to attain efficiency. A second tentative conclusion regards the potential of transfers between different transport modes. It seems that is only the marginal social cost pricing scheme that generate any revenues for the government.

3. Potential problems and inefficiencies in the European transport sector

In this chapter we give an overview of the European transport sector with a focus on potential efficiency problem and cost recovery. We discuss in turns roads, railways, airports and ports.

3.1 Roads

Most roads are owned by the public sector

In most countries, roads are a responsibility of the government. The responsibility is often shared between state and local governments. The main road network, which typically consists of motorways that bind together different parts of the national road network, is usually owned by the state. In general, it is also financed by the state. Local governments are often responsible for the local roads and also for financing. Not seldom, however, the state give various form of financial support to local road networks.

Government agencies are usually in charge of the national road network…

The state often delegates the operational responsibility for the main road networks to government agencies. Agencies make investment plans, promote traffic safety and a good environment, etc. The objective of the agency is decided by the government and is usually multidi-
dimensional, with both efficiency and equity dimensions. Profit maximization is not an objective. The agencies are in general financed by taxes. They receive annual budgets for investments, operation and maintenance, where the government decides how much of the budget that should go to each activity. Sometimes the agencies are allowed to finance certain investments with loans and toll revenues.

…but public companies exist
In Austria, a state-owned company, ASFINAG, is responsible for the main road network. In contrast to most government road agencies, ASFINAG is financed by fees paid by road users. The fees are a combination of tolls and vignettes. The fees should cover the company’s expenditures for maintaining and development of the Austrian motorways.

Public provision does not mean public production
Although roads are mainly supplied by the government it does not mean that the production of road services are carried out by the public sector. Traditionally, in the European road sector, public provision also meant public production. Both finance and labour resources came from the public sector. The last 10-15 years have witnessed a movement away from this organization form. More and more of the actual production have been contracted out to the private sector. The motive has been to promote cost efficiency. So far, the experiences are good. Cost savings in the range 10-20 percent have been achieved as a result of contracting out.

Private-public-partnerships are more and more common
A procurement form that seems to be more and more common in the road sector is public-private-partnerships (PPP). Here the private sector takes a much larger responsibility compared to traditional contracting. First, the private sector finances the construction of a road out of its own pocket. Second, the private contractor typically is responsible for both construction, operation and maintenance of the road for relatively long period, often 20-30 years. The private contractor’s revenue come either as annual payment from the government (shadow tolls) or from fees paid by the road users.

Road user charges
Revenues from the road sector consist both of taxes and fees and can be divided into the following categories:

- use of the road network (fuel taxes, tolls and vignettes)
- sales taxes
- sales taxes

Taxes on vehicle fuel are proportionate to road use and are therefore rather equitable. In Europe about two thirds of the revenues from the road sector comes from taxes on fuel, making it by far the single largest income from the road users. A secondary motive fuel for taxes is to restrain fuel consumption, and thereby correct some of the negative externalities caused by the road traffic. With regard to the second motive, the main drawback with fuel taxes is they not fully reflect the damage done to roads by heavy vehicles.

Vehicle taxes are used in many countries, most often in the form of yearly licence fees. The licence is usually differentiated with respects to types of vehicle size in order to reflect the

\[\text{UNITE}^{10}\]
costs that each type causes to the roads. In Europe, vehicle taxes constitute, on average, 17 percent of government revenues from road users.

Tolls are used for specific roads, bridges and tunnels. They charge directly for use of particular facilities. On average, toll revenues are about 5 percent of government revenues from the road sector. There is considerable variation, however, between countries. In France, Spain, Portugal and Italy toll revenues are 15, 9, 8 and 8 percent. In other countries, like Sweden, toll revenues constitute a very small part of total revenues.

Vignettes are fees that users have to pay for the use of an entire class of roads, typically motorways. Among the countries that have the vignette system, most charge only trucks. Two exceptions are Austria and Switzerland. On the average, revenues from vignettes only sums to about 1 percent.

Sales taxes on the purchase of new vehicles are levied by some European countries. On average sales taxes constitute slightly less than 10 percent of total revenues.

Revenue from the road sector varies between countries …
In Europe, governments revenues from the road sector, on average amounts to about 3 percent of GDP. There is considerable variation between countries, however. In France and Great-Britain revenues are over 4 percent of GDP, while in Norway and the Netherlands the corresponding figure is just over 2 percent. There are also considerable differences in revenue in absolute terms. In Luxembourg and Belgium, revenues from road users exceed 10 Euro per vehicle-kilometer, while in Greece revenues only are 3 Euro per vehicle kilometer.

… but in general by far exceed government expenditure on roads
As revealed by table 1 and 2 in appendix most European countries road user revenues are much larger than the amount of money government spend on construction roads and maintaining the existing ones. On average, the ratio between government spending on roads and road user revenues is .49. In other words, the revenues from road users are approximately double the infrastructure costs. There is, however, a large variation around this mean.

One extreme is Hungary, where spending is more than three times larger than revenue – the ratio is 3.22. Two other countries with high ratios are Austria and Switzerland: .89 and .90. These are both countries which relies to a considerable degree on user fees to finance the road network and that have lot of transit traffic in their road networks. France, Portugal, Spain, and Germany are also countries with ratios above or close to the mean. On the lower half, below the mean, we have most noticeable Denmark and Ireland with ratios of .09 and .11. Other countries with low ratios are Belgium, .25, Luxembourg, .26, and Finland, .30.

In sum, with one exception, Hungary, the road sector in Europe covers its costs, when by cost we mean, the governments expenditure on road infrastructure. As we will see below, because of externalities this cost concept is not the full or social costs of roads. But for the moment, we only consider infrastructure costs.

How are the surplus generated in the road sector used?
Some countries have ear-market road user revenues, that is, decided that revenues generated in the road sector should go back to the sector. Ear-marking occurs in Germany, Iceland, Switzerland, Austria, and in some of the new member states. But ear-marking is more an exception than a rule. In most countries, tax revenues generated in the road sector constitute
general revenues for the government. In fact, from a public finance perspective, it is a rather attractive form of taxation: fuel taxes are relatively easy to administer and collect, transparent, and cause relatively small distortions due to that demand for fuel is rather inelastic. Some countries, however, for example, Germany and Switzerland, seems to explicitly use revenues from road users to finance investments in other modes of transport, rail and waterways.

Most European countries have similar tax levels on fuel
The price of fuel does not vary much in Europe, it lay between 1.2 and 1.5 Euro. This means that taxes on fuel does not vary much either. In contrast, fuel taxes in the US and Asia are much lower than in Europe. On average, fuel taxes in Europe are about twice as high as the fuel taxes in the US. Interestingly enough, a much larger proportion of the revenues generated by the fuel tax go to the road sector in the US than in Europe. In the US, about 75 percent of the revenues from fuel taxation return to the road sector. The corresponding figure for Europe is much lower (Queiroz, (2003)).

Public provision of roads is justified by markets failures…
The governments role in the provision of roads can to a large extent be explained in terms of markets failures: monopoly power, externalities and public goods.

First, production of road services is associated with significant returns to scale, with decreasing average costs over a large output range. Hence, competition between different road providers in the same region is not desirable. A monopoly provider is necessary to achieve technical cost efficiency. The risk for abuse of monopoly power is dealt with by government ownership, usually in the forms of government agencies with not-for-profit objectives.

Second, many roads may be thought of as pure public goods. The marginal cost of having one more vehicle on an uncongested road is close to zero, at least for a passenger car. For a heavy vehicle, the marginal cost is larger, which motives that they should pay a fee equal to this cost. In addition, in many parts of the road network there is, or has at least been, problems charging a vehicle for use of the road. Thus, public provision of road services free of charge for passenger cars seems like a rational arrangement from a microeconomic perspective. However, other parts of the road network do not meet the public good criteria. The marginal cost of having one more vehicle is clearly not zero in a congested road because of externalities. Furthermore for more and more parts of the road network there is no longer technical difficulties in charging vehicles for use of the roads. In other words, exclusion is now easier than what it has been in the past.

Third, roads are associated with major externalities. The externalities are positive and negative. Usually, people tend to dismiss the idea that road transport bring along any positive externalities other than those visible on the road, and hence internalised by road users. For instance this seems to be the position of many transport economists and also the European Commission. This position has, however, recently been criticized by Baum et al. (2008) in a critical review of the EC internalisation policy. They argue that road transport gives rise to a number of external benefits: mobility improves the division of labour, increase productivity and leads to more growth, higher incomes and employment. These effects, they mean, are entirely neglected by the EC as expressed by so called CE Delft Handbook.

We conclude that their critique has some bearing. It seems likely that there are positive externalities with having a good road system. A society in which individuals and firms can transport by car, bus or truck in a fast, reliable, well-developed, and maintained network of roads,
almost certainly function and develop much better than a society in which the road network is of poor and unsufficient quality and quantity. Because the private return to building and maintaining a road network is likely to be much smaller than the social return, private provision would result in an insufficient capacity. This is probably one of the main reasons for public provision of roads.

The negative effects of roads are more familiar. Both production and consumption of roads give rise to various negative externalities. Pollution, congestion, accidents and noise are all well known externalities that lead to excessive driving and hence needs correction. These are all external effects that the EC support and recognize. Also regarding the negative externalities from the road sector there has been some recent critique (Baum et al. (2008)). They argue that the EC in an incorrect way count as externalities that need corrective measures costs that are already internalized. For example, congestion costs have reciprocal nature, which means that they are borne by the motorists who bear them. Hence, they are already internalized. Accident cost is another example. About two thirds of accident costs are covered by insurance companies, which mean that these costs are already reflected in the insurance premium motorist pay. Thus, the external costs that EC argue should be corrected are too large.

… and equity reasons
Improving economic efficiency is not the only motive for public provision of roads. As pointed out in section 2.3, an important policy for many governments is the pursuit of equity. In practice, this has several implications. For example, tolls create economic inefficiencies when they are levied on roads, tunnels or bridges, where capacity is sufficient so congestion is not an issue. Tolls can in these situations be motivated by equity reasons. Those who benefit, the users, are also those who pay, which might be seen as only fair. Some efficiency is here traded for a more fair or equitable solution.

From a theoretical perspective current practice seem fairly efficient
Looking upon the current practice used in Europe it is not clear that there are any obvious inefficiencies. The public good characteristics of many parts of the road network calls for public provision and zero price of usage. This means that some kind of tax financing is necessary. As we saw above, road users pay several different forms of taxes.

Vehicle licences may be thought of as the fixed part of a two-part tariff pricing scheme with a zero marginal fee – a fairly efficient practice.

The negative externalities are dealt with through market based Pigovian taxes and subsidies and direct regulation. The fuel tax is in principle a rather efficient way to deal with the negative external effects in the form of pollution. Whether current taxes are at the correct level is more doubtful, as reflected by the discussion above. We soon return to this issue. Some governments, subsidize “environmental” friendly cars that have a lower fuel consumption. Governments also correct the externality problem through direct regulation, e.g., on emissions, safety requirements etc.

One major inefficiency, however, seems to be the severe congestion problems that are in many large cities. These are a result of a combination of “free access” and limited capacity. Some cities, like Stockholm and London, have now introduced congestion fees for entering the cities with motor vehicles. The effect of the fees has been to decrease traffic and thereby congestion. By differentiating the fees depending on the demand – higher fees during peak
hours and lower fees during off-peak hours – the governments are actually using a form of peak-load pricing.

**Does the road sector really cover its costs…**

One issue often debated is whether the road sector covers its costs or not. According to some people it is clear that the road sector is overcharged and by far cover its costs. Other people suggest that it is not true that the sector is overcharged when one take into account the external costs caused by the road users. To shed light on this issue, we again turn to figures calculated by the UNITE-project (Table 1 and 2). According to that study, on average, the road sector does not cover its total costs (defined as infrastructure costs, congestions costs, air pollution costs, noise costs, costs of global warming, and external costs of global warming). The ratio between total revenues from road users and total costs is .76. Hence, on a European scale, if total costs generated by the road sector were to only be covered, taxes and fees need to go up with approximately 25 percent. Again, variation among countries is large. For example, in Austria, France, Germany and England, the road sector does not cover the total social costs. In Belgium, Denmark and Sweden, on the other hand, revenues do cover the social costs.

What conclusions can we draw from these numbers? First, we must note that these numbers are quiet old, from 1998, and things may have changed since then. Second, one may question whether all relevant externalities are calculated – remember the discussion above. In other words, some caution is warranted. But suppose that these numbers were accurate. What does it tells us?

A first conclusion is the effect of an implementation of marginal social cost pricing would not be uniform. For these countries where the road sector does not cover its costs, the gap between revenues and total costs indicates that the negative externalities are not fully internalised. There is too much driving with the wrong sorts of cars at wrong times and places. For the countries where revenues do cover costs, the opposite is true. From an efficiency perspective, in these countries the car sector as a whole is overcharged.

A second conclusion is that in the former group of countries, implementing marginal social cost pricing would mean tax rises, while in the latter group it would mean lowering road use related taxes.

**…and does it matter?**

It seems that the degree of cost recovery in the road sector vary quiet substantially between countries. In some countries, more than full cost recovery is achieved, in other it is not. Under certain conditions this can be interpreted as an indication of that externalities are handled in a non-optimal way. When externalities are dealt with by Pigouvian taxes, efficiency is achieved when full cost recovery – not more, not less - is achieved. So if one interpret current practice in the road sector (taxes on fuel, etc) as a mainly Pigouvian method of dealing with the externalities, then the result above, are meaningful and give clear policy recommendation: raise user taxes in countries like France and Germany, and lower them in countries like Sweden and Denmark.

However, as noted in end of the last chapter, Pigouvian taxes are only one of several methods of handling problems with externalities: regulation and marketable permits are two other. If these methods are used instead, then cost recovery is not necessarily a very interesting measure of performance.
More price-discrimination is likely to both increase revenues and improve efficiency but at the cost of equity

Economic theory suggests that price discrimination usually improves efficiency compared to a single-price policy in areas where the supplier has market power (see section 2.). Price discrimination could either be based on willingness to pay or on cost of supply. Current European practice clearly has elements of price-discrimination. But by going further is it very likely that both efficiency and revenue would increase, at least for the parts of the national road networks that have severe congestion problems. Some cities have already introduced congestion charging. This is an example of user fees that are suitable to extensive price discrimination. Greater variation in congestion charging is likely to have the double positive effects on increasing revenues and reducing troublesome externality problems. This is positive. The drawback is, however, as nearly always in political decision-making less that increased efficiency will have a price in terms of greater inequity.

3.2 Railways

European railways have over time lost market shares to road transports

The last 30 years, the European railways have continually been losing market shares. In 1970 rail had a market share of nearly 22 percent of the European transport market. The corresponding figure for year 2001 was 7.8 percent, a significant decline (Pelkmans et al. (2004)). It is mainly road transports that have gained markets at the expense of the railways. During the period 1990-2001 road transports increased by 35 percent, while rail freight transports decreased by 6 percent. The situation is somewhat worse for the passenger transport by rail.

The railways decline is due to a mix of exogenous and endogenous factors

How can the decline for the European railways sector be explained? Pelkmans et al. (2004) point to whole number of factors, which can be divided into exogenous and endogenous factors. One significant factor in the first category is the transformation of the business and industry landscape that has happened in Europe, and the rest of the industrial world, the last two or three decades: a movement from large-stock-based to just-in-time production processes and from low-value/high-volume to high-value/small-volume products. This transformation had required a different, more flexible transport system than what the railway has been able to offer. Hence, the increased market share for the more flexible road transport system. Another significant exogenous reason is policies and investments that have favoured road transport instead of rail.

But exogenous reasons are not the only reasons. Pelkmans et al. list a number of failures, mistakes and shortcomings that stem from the railway sector itself. The railway sector has to a large extent failed to respond to the changing market environment and the consumers demand. For example, while the trend has been towards cross-borders transports, the railway industry has mainly remained a domestic industry. Lack of competitive pressures and fragmentisation of the European rail industry are other endogenous factors that have contributed to the decline.

The situation for the European railway sector seems to have improved the last years

The last couple of years have witnessed sign of a more positive development for the European railway sector. In several countries, high speed trains have been successful in reaping market shares from domestic air transport. Sharply raising fuel prices, making both road and air transport costlier, is another factor that benefits the rail sector. More and more passengers are
substituting road air and transports for rail transports. Concerns for global warming also work in favour of the rail sector. Urban commuter transports is another sector with a successively growing demand for rail travel. In some countries, there are indications of a shortage of rail capacity to meet the increasing demand.

For a long time network management and train operation where vertically integrated in Europe
In Europe, the railway sector where for a long time organized as vertically integrated national monopolies. Both infrastructure management and operation of trains were typically carried out within the same entity. This organisation was deemed suitable given the characteristic features of the railway sector: large economics of scale and scope, a technically demanding and complex system, and need for centralisation due to the systems highly connected character. However, it seems fair to say that these vertically integrated public train monopolies over time became cost inefficient and quite expensive for the governments and suffered from a lack of consumer perspective.

The EU has worked for liberalisation of the European railway sector since the early 1990
Mainly as a response to the perceived underperformance of the rail sector, the European Union in the beginning of the 1990s began a reformation of sector. A three stage approach of regulation and liberalisation was adopted. The first stage was a Directive on the accounting separation between infrastructure and operations. The second stage, the First Railway Package, was an attempt to liberalise the freight sector through the introduction of open access and forms of competition. It also contained a number of technical measures to remove various entry barriers. The third stage was launched in form a Second Railway Package, a promotion of harmonization of safety requirements and certifications.

Several but far from all member countries how unbundled infrastructure management from traffic operation
One route to liberalisation that several countries, for example Sweden, Norway, Denmark, the Netherlands and England, in Europe have taken is vertical separation of the rail infrastructure from the train operation. In all these countries but England, the responsibility for the infrastructure has remained in the hands of a public organisation, while attempts have been made to open competition in the operation of trains. The basic idea is that economics of scale and scope mainly exist in former activity. Several countries have, however, kept the model with vertical integration: Germany, Austria and Italy. Also vertical integration is still the dominant model in other parts of the world. For example, Amtrak, the big American train operator retains control over the rail infrastructure, as the Japanese company the Japan Rail Freight organisation.

Sweden was one of the first countries who unbundled network management from train operation
In 1988, the Swedish National Railway, SJ, was vertically separated into one infrastructure provider, Banverket, and one train operator, SJ. Banverket has from the start been financed with general taxes and receives an annual grant. Its objective is to promote the development of the Swedish railway sector and to treat all operators on equal terms. SJ receive no tax support and must earn its own revenues by producing and selling train transport to firms and passengers. It needs to pay fees to Banverket for using the rail network. These fees should correspond to the short run marginal cost. At first SJ operated both freight and passenger traffic. To date, the freight business is carried out in an other company, Green Cargo. Both SJ and Green
Cargo are still state-owned companies. For a long time SJ has had a monopoly on running train operations on most of the national network. Its objective has since the start been to be as profitable as possible. However, it is only in the last couple of years that is has been commercially successful and able to meet the owners stated requirement on rate of return. In other words, despite a seemingly favourable market position and paying the short run marginal cost for rail usage, SJ has had a hard time making a profit.

**In the United Kingdom, most of the railway sector is the hands of private firms**
The railway sector in the United Kingdom (UK) is the railway sector in Europe that has gone furthest towards privatization. In the UK, vertical separation between rail infrastructure and train services has been in place for about 15 years. The former vertically integrated public monopoly train company, British Rail, was split in a number of firms for train operation, maintenance of the rail infrastructure, vehicle provision and maintenance. A still unique feature of the British attempt to reform the rail sector is the privatization of the infrastructure provider with the creation of Railtrack. The idea was that Railtrack should be run on a solely commercial basis which revenues coming from its customers, the rail operators. The system, however, failed after a couple of years and Railtrack was brought back public ownership in the form of public company, Network Rail, that partly receives its revenues from taxes. Today there are about 25 train passenger service operators and a number freight operating companies. The former category of firms is granted franchise contracts for 25 specific routes after a competitive tendering process. Despite, the failure of Railtrack, many people judge the reformation in England as mainly positive.

**Major subsidies goes to the railway**
The European railway sector is heavily subsidied from public budgets (Pelkmans et al. (2004)). It is estimated that the railway sector subsidies are second only to expenditures related to agricultural policy. The subsidies can be divided into three categories: financial and miscellaneous, payments for investments and payments for services and maintenance.

A study conducted by the European Commission reveal large variations in subsidy levels both within these categories and between different countries. According to this study, the three countries with the smallest subsidies (measured as payments – Euro cents per traffic unit) are Portugal, Finland and Sweden, with subsidies amounting to 2,3 and 3,5 Euro cents per traffic unit. There are, however, large differences between these three countries as to what these payments go to. Portugal only subsidy investments, while Finland mainly finances payments for service and maintenance. In Sweden, subsidies, are about evenly divided between investments and service/maintenance.

Medium subsidy countries are Austria, Italy, Belgium, Denmark, the Netherlands and Ireland, with payments ranging between 8 Euro cents per traffic unit to 16 Euro cents per traffic unit. Also in this group variations are considerable regarding the activity being subsidied. In this group the most substantial subsidies goes to financial support.

The largest spenders with public money to the railways are Luxembourg and Greece. The latter country has by far the largest subsidies, over 30 Eurocents per traffic unit.

In sum, the national railways in Europe receive substantial subsidies for the provision of services.
The reliance of user fees to finance the infrastructure vary considerable between countries
The degree to which fees charged by rail operators finance the rail infrastructure vary substantially between the European countries. In several countries user fees cover over 60 percent of the infrastructure cost: Denmark (60%), France (63%), Germany (60%), Great Britain (50-100). In other countries, cost recovery is much smaller. In Sweden and the Netherlands, cost recovery is 5-10 percent and 12 percent.

The rail-industry is a network industry characterised by significant economics of scale of scope
The rail transport system is a so called network industry. It is technically complex and has a rather extreme cost structure with a high share fixed costs and low marginal costs: fixed costs for infrastructure account on average for approximately 90 percent of total infrastructure costs (Campos & Cantos, (2000)). This gives rise to considerable economics of scale, both regarding the infrastructure and the train operation.

Rail service provision, at least in Europe, is also characterised by its multi-product nature. There are two main products: freight services and passengers services. Within these two categories there exists a spectrum of sub-services, especially for freight. Passenger services have two major product segments: long-distance services and commuter line services. This multi-product nature gives rise to considerable economics of scope.

Finally, the network structure give rise to network effects that increase the value of services the more interconnections that can be guaranteed in the network. It also allows a certain degree of cross-subsidiation of routes to occur.

Together these features mean that the room for many actors, and thus competition, is limited relative to the other transport systems. In addition, is a strongly connected system (see section 2.1), which require a high degree of centralisation.

Numerous technical and legal barriers to entry
One factor that contributes to the limited amount of competition that characterizes the European railway industry is the existence of numerous technical and legal barriers to entry (Pelkmans et al.):

- a fragmented licence scheme
- lack of interoperability
- restrictive train drivers licenses
- lack of a second hand market for locomotives and an absence of leasing
- priority allocation and safety systems
- hindrances to cabotage

According to Pelkmans et. al these entry barriers must be removed before the competitive situation can be improved.

National giants dominate both freight and passenger traffic…
In several countries, the operation of trains is dominated by former national monopolies. This is true for both passenger and freight. Competition is, however, gradually opened more and more. New firms have been started that challenge the incumbents. The third railway package that includes passenger service liberalisation is likely to further reinforce these tendencies.
...but their market power are limited
The circumstance that competition on the tracks is still limited and markets are dominated by large public companies does not mean that these firms have large market power. Both the passenger and freight traffic face tough competition from other modes of transport, most noticeable from buses and cars. Furthermore, several countries have introduced competitive tendering procedures for parts or the whole rail network. Finally, and related to the first point, the demand for railservices is likely to be rather price sensitive, which limits market power. It is illustrative that SJ, alleged to be one most efficient train operators in Europe, have trouble covering costs.

Problems with externalities are relatively small in railway sector
The external costs associated with the rail transport industry in Europe are very small in comparison to the other modes. Table 3 show that the calculated external costs (air pollution costs, noise costs, costs of global warming, external costs of accidents) for the railway sector very between 1 percent to 10 percent of the infrastructure costs. Thus, the need for corrective measures is not a major issue for the sector.

Should cost recovery be strived for?
As we saw above, the variation in cost recovery of the rail infrastructure is substantial between countries in Europe. The ambition for some of these countries is as far as possible to reach full cost coverage, that is, railway companies should pay fees sufficiently high to cover the cost of the infrastructure. Is this a wise policy? From an efficiency point of view is does not seem so. The marginal cost constitutes only a minor share, around 10 percent, of the total infrastructure cost. Given the relatively small external cost, an efficient marginal social cost pricing, imply user charges in the range 10-20 percent of the infrastructure costs. This pricing would promote an efficient use of the existing rail network. Average cost pricing would mean very much higher fees. Given the relatively price elastic demand for rail services, this is likely to lead decrease demand for rail services. As a result, the rail infrastructure would be inefficiently little used. In other words, cost recovery should not be strived for.

3.3 Airports
Travel by air has increased enormously over time
In a historical perspective, the air transport sector is a relatively new mode of transport. It is, however, a sector that has grown enormously over time, especially the last couple of decades. In Great-Britain, the number of passengers passing the British airports increased from about 100 million passengers in 1991 to nearly 230 million passengers in 2006 – 130 percents increase. A recent forecast estimated a further large increase to between 400 million and 600 million passengers by 2030. Similar developments are seen and expected in other European countries. If the current trend with a successively higher price of oil and other measures to reduce emissions of green-house gases continue this expected development may dampen.

The airline industry is characterized by fierce competition and low profits
“What is the surest way of becoming a millionaire? Being a billionaire and starting an airline?” This joke captures in a rather brutal way the economic realities of the airline industry. It is an industry where the internal rivalry is very strong and the average profitability low, even in good economic times. The liberalisation of the industry that occurred in the 80s and 90s has increased competition, not least due to low the entry of low-cost carriers like Ryan Air and Easy Jet. The fact that the industry is so competitive can be explained by basic microeco-
nomic theory: a large number of competing airlines, a highly homogenous product, and relatively low barriers to entry. These forces all affect work to increase competition.

**Low-cost carriers have transformed the industry…**
One of the most significant changes that has taken place in the air transport industry in the last 10-15 years is the entry of low-cost carriers. It all started with a relatively small and regionally based American airline, Southwestern Airlines, who brought a new strategy concept to the industry in the early 1980s. The core of the strategy was to dramatically reduce costs and increase productivity. Southwestern Airlines has been profitable every year since the start, which is unique to the air transport industry. Beginning in the early 1990s, several European airlines, like Ryan Air and Easy Jet, adopted the American firm’s concept. By cutting costs in all parts of the value chain, the low-cost carriers have been able to offer passengers very attractive prices which has increased demand for their services. At the same time, they have given the established airlines a tough time.

**…partly by avoiding the expensive major airports**
An important part of the low-cost carriers strategy have been to fly from smaller airports, often located outside the largest cities, which host the major airports. By doing this, they have been able to avoid the high fees charged by the major airports. The lower fees have been translated into lower costs and a competitive advantage.

**The air transport industry is currently going through a major cost crisis**
The dramatic increase in the price of oil that has occurred the last couple of years put a heavy burden on the air transport industry. With oil prices at the current level, the average airline is going to lose money, unless they can pass on the increase in cost to passengers. Given the competitive forces that characterize the industry this is not likely to happen. Other cost drivers are increased security measures and charges from airports. Measures to deal with the negative externalities caused by air transport is another factor that may cause costs to rise.

**Domestic air travels face tough competition from high speed trains and currently losing market shares**
Several countries in Europe have invested in high-speed rail connections between major cities, for example France, Germany and Spain. In those countries, high speed trains have become a tough competitor to the domestic air transport industry. Also in other countries, which do not have, have speed-lines, it can be seen that the air transports face stiffer competition from trains. This trend is likely to continue. One effect of this is that capacity will be freed at airports which can be used for the international air transport traffic.

**The ownership structure of the European airports vary a lot between countries**
It is possible to distinguish at least five different ownership structures of the European airports.

First, we have a number of airports that are owned by the state. For this category of state-owned airports the main objective is not profitability. The goal is usually broader with social and regional dimensions to it. This does not, however, that the commercial side is totally absent. The state often demands a required return on capital.

The second group of airports is also owned by the state but is more commercially oriented and independent from the owner. This second group is typically run as public companies.
A third group of airports is owned by regional interest, like municipalities or cities, can be found in for example Germany, Denmark, England and France. To promote regional interest is an important objective for these types of airports.

A fourth category is airports with a mix of private and public ownership. The cooperation is often the result of need for capital for investment that has required private participation. The reason that state keep on ownership is to have guarantee of certain influence.

The fifth category is airports that are solely in the hands of the private sector. Most major airports in Great-Britain belong to this category. The largest airport in the Nordic countries, Kastrup, is since 2005 owned by a majority of private interests. Today, Kastrup is one of the most profitable airports in Europe.

There has been a trend towards privatization of airports...
The air transport industry in Europe was for long time dominated by public airlines and public airport. Over time this has changed. A privatization wave started in England during the 1980s, where first the state-owned airline, British Airlines, became privatized and later on the major airports. In the end 1990s, extensive privatization plans for airport was made in Germany and France. To date, these plans have not yet been realized.

...largely driven by cost-inefficiencies and investment needs
The main motives for the privatisation plans of major public airports have been a combination of a perceived low cost inefficiency and lack of money for investments. With privatisation comes more commercially oriented interest in to the picture something regarded as a positive injection for cost-efficiency. Capital for investments is the other main motive for bringing in private interests. Large investments in new infrastructure with more capacity have been necessary to meet the expansion of the air traffic that has occurred the last 10-15 years. In several countries, the state saw private capital as way to free the burden of making these large investments themselves.

Airports are attractive investments
It has turned out that large airport with a lot of traffic are attractive investments for investments bank and companies. The major airports offer a stable business with increasing revenues and relatively small risks. An indication of the attractiveness is that the investment company Macquarie Airports paid more than 30 percent more the existing stock price when they bought Kastrup.

In most countries, the large airports cross-subsidize smaller airports
Far from all airports in Europe is profitable. Regional airports, with relatively little traffic, have a hard time breaking even. Revenues from landing fees and commercial activities are much bleaker than for the major airports with their large and stable demand. Many countries, however, believe that they need a functioning national air transport system, not least for regional-development reasons. For this reasons, several countries have established a system where the major, profitable airports use part of their revenues to subsidize the operation at smaller airports. These transfers-systems are debated and are from time to time overseen.

In general airports are financed by fees and charges, not general taxes
In contrast, to both rail and road transports, taxes are in general not used to finance investments and operation of the airports. Instead, airports are mainly financed by airport fees paid by the airlines. The structure of these fees should follow the guidelines established by the In-
ternational Civil Aviation Organization. This means that the fees should be non-discriminatory, cost-based and transparent.

Fees are in general proportional to the size of the aircrafts and the number of landings. The fees consist of several components: a passenger fee, a starting fee, a fee for flight control, a fee for emissions and noise, etc.

The airports are also charging fees for security. After the terrorist attack in the United States in September 2001, the European Union sharply increased security at the airports, which increased costs. To finance these costs a new fee was introduced. It is a fee per passenger that the authorities charge the airlines.

The size of fees are not regulated. This means that there exist considerable variations between different member countries. The major French airports is generally believed to have the highest fees, something claimed to be a result of a low cost inefficiency.

**Non-flight related revenues increase in importance**

Revenues from commercial activities like shops, conference services, parking spaces etc, also constitute a major source of revenue, in particular for the major airports.

**The major airports in each country probably have considerable market power**

Most airports have a monopoly situation, in particular the major airports. It is true that they to some extent face competition from local or regional airports. Especially, they feel the competition from smaller airports, who host the major low-cost carriers. But it is likely, that the major airports have a considerable market power. Demand for their services is very strong and they offer services that have no good substitutes. The fact that they are profitable and attractive investments are indications of their market power. The fierce competition that exists downstream, among the airlines, also means that most of the bargaining power lay with the airports. The airline industry complains over high fees and high costs. These are also indications that the major airports have market power that they use.

The circumstance, that the demand for the major airports services is relatively inelastic means that the loss of allocative efficiency is not as great at it would be with a more elatic demand. In other words, it is not likely that the amount of flying has decreased by much as consequence of market power. Instead, the consequence is an income transfer from airlines and passengers to airports.

At the same time, it is probably the case that most of the major airports are currently not fully exploiting their market power. It might well be the case that a further privatization of the large European airports, with a stronger business focus as a result, would lead to a higher exploitation of market power. The fact that Kastrup is one of the most profitable airports in Europe is probably a combination of a more extensive exploration of market power and higher cost efficiency.

Many smaller airports do not have market power. They face a small demand and have trouble cover their own costs. In fact, without the cross-subsidation from the larger airports and other financial support from governments they would not be able to survive.

**Travelling by air gives rise to negative externalities that currently is not fully internalized**
Air transports give rise to several negative externalities, mainly in form of pollution and noise. Lately, congestion in the air has increased, which is another externality problem. As mentioned, the current fee structure contains exhaust and noise components. These fees may be regarded as Pigouvian taxes – attempts to force airlines to internalize these externalities. The last couple of years there have been discussion about the air transports industries contribution to global warming. Aircrafts go on oil-based fuel and emits greenhouse-gases. Estimates suggest that globally this industry stand for about 3-5 percent of the total emissions of greenhouse gases. Currently, these emissions are not the subject of any regulation or other corrective measures. There are plans, however, to introduce marketable permit system for CO2 in the European air transport sector.

The European air industry does not cover its costs
The UNITE-project calculated total costs and revenues for the European air transport sector (Table 5 and 6). Again, the numbers are from 1998, so they may not be very accurate, especially given the large growth in air transport that has occurred the last 10 years. These estimates suggest, however, that on average, the industry is not covers its costs. The main external cost seems to be costs of global warming, underscoring the importance of dealing with this externality.

Should airports be price regulated?
The circumstance that the major airports are likely to have considerable market power, due to the combination of lack of competition and relatively inelastic demand, raise the issue of whether a stronger regulation of the fees that airports are allowed to charge is wanted. As long as the airports stay in hands of the public sector the answer is probably no. Although airports are getting more and more commercially oriented, a public ownership is a guarantee that the profit motive does not grow too strong and overshadow other interests, like regional development etc. In case of full scale privatization, things change. Then some kind of price regulation is likely to be needed to prevent the market power from being exploited too much.

The airport industry must deal better with the negative externalities it create
The airport transport industry is not yet internalising the externalities due to emissions of greenhouse-gases. The industry argues that a system with marketable permits is not necessary given the current development of the price of oil. They mean that airlines are now forced to investment in more fuel-efficient aircraft to cut costs. Airlines with old, less fuel-efficient aircrafts will not be able to survive. In other words, markets forces will self handle the externality problem. Although, this reasoning is based on sound economic reasoning, it is probably not sufficient. It ignores the substitution effect that occurs when planes get more fuel efficient and emit less greenhouse gases – traffic level will not be constant but increase. Hence, some kind of public intervention is needed. A marketable permit system have several advantages.

3.4 Ports
Ports have a major role in the world trade system
For centuries ports have been economic units of major importance in international trade. The important role of ports has not decreased, quite the opposite is true. Today nearly 90 percent of goods exchanged in the international trade process in the world rely on maritime transport. Port traffic is currently growing at an annual rate of three percent. This development means that well functioning ports are import for governments, in Europe as well as in other parts of the world.
A common European port policy does not yet exists
In comparison to other transport modes, ports stands out in the sense that a common European policy has not yet emerged. Both member states and the major ports have resisted the launching of a common port policy, seemingly out of fear that there autonomy would be reduced and the wide variety of political objectives, financial structures and patterns of ownerships that exists among European ports (Trujillo, et al (2007)).

Port authorities are typically public institutions
Usually, ports offer a large multiplicity of services that may be divided into three main categories: infrastructure services, vessel related services and cargo related services. These various services are often carried out be different firms. This great multiplicity off services and suppliers create a need for a coordinating actor. In most ports, this function is carried out by a public organisation, the port authority.

There exist three main organizational models for ports
One may divide the European ports into the three different organizational models (Trullijo (2007)). The models differ mainly in the respective roles for the port authorities and the various private suppliers.

- The Landlort port. In this model port authorities own and manage the port infrastructure. Private firms provide port services and own their assets – the superstructure (building etc.) and equipment (cranes).
- The Tool port. In this model, port authorities owns the infrastructure, superstructure and equipment. Private firms are in this model mere providers of services and are renting port assets through concessions or licences.
- The Services Port. In this third model, port authorities are responsible for the port as whole. Port authorities own both the infra- and superstructures, and equipment. They also hire employees to provide services directly.

Substantial diversity with regards to financing and charging practices
One of the main observations from the ATENCO study11, was the great diversity prevailing among ports in Europe regarding financing and charging practices. This diversity is a result of a number of different factors, like various judicial and cultural traditions, divergent port management styles, degree of autonomy etc. Two main approaches can be identified: the Anglo-Saxon approach and the Continental approach.

In the countries that have adopted the Anglo-Saxon approach ports tend to have a clear commercial orientation in which users (mainly shippers) bear all costs generated in the production of various services provided by the ports. According to this doctrine ports should be profitable or at least generate revenues sufficiently to cover costs. In other words, no tax subsidies should be necessary.

The continental approach is characterized by less a commercial orientation. Instead, in this view, ports are considered as being part of the society’s psychical infrastructure, much like roads and railways, for which the public sector typically has taken on a responsibility to provide and contribute to financially. According to the continental approach ports are vital in

contributing to economic and regional development and the government is therefore willing to promote ports in fulfilling this role. Full cost recovery, in the sense that user charges cover costs, is not necessarily strived for and public subsidies are often a mean to cost recovery.

These differences in the views of financing and subsidisation of ports, of course, complicate the task of finding uniform methods of cost recovery.

**Port pricing is very complex as indicated by both theory and practise**

The task of finding uniform pricing methods is also complicated by the fact that port in itself pricing is a very complex issue. Both the academic literature and practice witness that. There exists a considerable amount of academic work on port pricing (see for example, Haralambides et al. (2001), Marlow and Pettersen &Strandenes (1999)). The main pricing principles discussed in the literature are: cost-based pricing, methods for cost recovery, congestion pricing and strategic port pricing. For each principle, a number of price systems are suggested.

Regarding the first category, long-run marginal costs is deemed to be the most appropriate cost basis if economic efficiency is the main policy objective. Since economics of scale generally exist in provision of port services, this pricing rule is likely to require subsidies for ports to cover the total costs. Fixed costs constitute a significant share of total costs, figures in the range 60-80 percent is often mentioned. Variants of Ramsey pricing schemes and two-part tariffs are often discussed when cost recovery is on the agenda. It seems fair to say that no clear-cut pricing scheme is dominating other alternatives.

Port pricing in practise is even more complex. One observation is that there are elements of cost recovery in traditional port pricing, which either reflect average cost pricing or a combination of charges and subsidies (Strandenes et al.) Also it turns out that approaches based on average costs appear to perform reasonable well in approximating marginal costs (Robinsson (1991)). Second, port pricing tends to be rather non-transparent. Third, a best practice port pricing formula does not exist (Haralambides et al.). They report a case study of three ports in the UK that apply a pricing policy aiming at full cost recovery. One main finding was that these three ports used a variety of pricing principles simultaneously. In sum, the issue of infrastructure pricing of ports is complex.

**Public support of ports is motivated by a number of factors but research indicate that cost recovery is both possible and desirable**

According to Trujillo (2007), there are four main justifications for the practice of financing port infrastructure with tax money:

- **The strategic role** that ports have in many countries. Because both export and import of goods to such a large extent rely on maritime transports, ports have a strategic role in the value chain. A policy that makes a regions main port less attractive for international traffic, for instance by charger higher prices than competing ports, is seen as risky, and vice versa.
- **Ports stimulate economic development.** As already mentioned, well functioning and attractive ports are seen as important tools for economic growth.
- **Market failures.** As with other physical infrastructure, basic port infrastructure is indivisible, requires large sunk costs, have a long life time, and are constructed in a spe-
specific space for a specific use. This means that there exist economics of scale, which in turn may create something of natural monopoly.

- Some port services have a nature of *public service obligations*

Trujillo argue, however, that none of these four circumstances, even in the case they are valid, is incompatible with strategy to recover costs by way of a well-designed tariff system. This view is supported by the main conclusion of a comprehensive academic literature review in port pricing undertaken within the ATENCO project: port pricing can and should be based on costs (Haralambides, 1991 et al.)

**Ports themselves seem to be in favour of cost recovery**

A survey conducted within the ATENCO-project involving 13 port authorities and port users in Europe indicated a broad support for the adoption of overall full cost recovery. A majority of the ports supported the adoption of user pays principles. Moreover, most port authorities expected that that an implementation of a full cost recovery pricing would have little effect on price level. Haralambides et al. suggest, however, that this view, although true for some ports, will not hold true for others. For example, an introduction of a full cost recovery requirement in the Mediterranean ports, was viewed as infeasible due to the fierce competition among ports in this region.

**Market power seems limited for most ports**

As mentioned, there exists economics of scale in production of port services: economics of scale exist in port infrastructure, for instance in aids to navigation and cargo storage, and in cranes for loading and discharging cargo. In order to exploit these, there typically only exists one main port per city. In principle, this could give ports monopoly power. However, it seems that there is a strong rivalry between ports, at least in parts of Europe, which puts pressures on prices. Hence, market power may not be a big problem for ports.

**Negative externalities in form of emissions are an increasing problem**

Maritime transports use mainly oil-based products as an energy source, much like air transports. This means that the transports give rise to various emissions harmful for the environment and the health of people and animals. The international maritime industry is today one of the largest polluters of certain emissions like sulphur. Given that the current trend continue with more and more maritime transports, emissions are likely to increase further. Until only recently, virtually no measures to mitigate these emissions existed. This spring, however, the International Maritime Organisations member countries agreed on plan to sharply reduce emissions. If this plan is effective, the externality problem will be smaller in the future.

### 4. Concluding remarks

This paper has studied various aspects of marginal social cost pricing and cost recovery in the European sector. In this last section, we make a few concluding remarks.

**Efficiency problems exists in all transport modes**

Not surprisingly, efficiency problems exist in all four transport modes studied in this paper. From a microeconomic perspective, the characteristics of provision of transport infrastructure and use of it: large economics of scale, negative externalities and public goods, means that inefficiencies can be expected. The fact that governments in most countries play such a prominent role in infrastructure provision can to a large extent be explained by these potential
market failures. One finding that comes out of this study is that the main efficiency problems differ between transport modes.

For roads, the main economic inefficiency seems to be the negative externalities that materialize in the form of pollutants and congestion.

For railroads, X-inefficiency seems to be the main problem. The railway sector lacks in both cost efficiency and dynamic efficiency, circumstances that reduce the rails ability to compete against other modes of transports.

For air transports, market power and negative externalities are the big problems. The major airports in Europe are likely to have considerable market power, which is one factor that contributes to the cost crises of the airlines – high user charges. The air transport sector also give rise to too inefficiently high amounts of emissions harmful for health and contributing to global warming.

For the maritime industry and the European ports, negative externalities in the form of pollution seem to be the major inefficiency problem.

**Cost recovery is not a virtue in its own right**

Sometimes people seem to argue that cost recovery is good thing, something to strive for. From a public finance perspective one may question this view: cost recovery is not a very interesting issue. First, all cost must be covered, with user, taxes or something else, which means from one point of view the concept has no meaning. Second, there is no good economic argument for why a specific activity carried by the public sector should cover its costs. Third, in case external costs, the issue of cost recovery seems to be confused with the issue of correcting externalities. It is the externality problem that it is of interest to reduce and not to make the party causing the externality to bear its costs. It is true that in the case a Pigouvian tax is used to correct a negative externality the amount paid is equal to the external cost, but that need not be the case when other measures to deal with the externality, like regulation or a marketable permit systems, are used.
Annex to Appendix A

This appendix contains a number of tables over costs and revenues for different transport modes calculated in the UNITE-project. All number are from 1998.

Table 1
Costs of road transport for European countries in 1998, in Million Euros

<table>
<thead>
<tr>
<th>Country</th>
<th>Infrastructure Costs</th>
<th>Congestion Costs</th>
<th>Air Pollution Costs</th>
<th>Noise Costs</th>
<th>Costs of Global Warming</th>
<th>External Costs of Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>4382</td>
<td>1555</td>
<td>833</td>
<td>329</td>
<td>36</td>
<td>1367</td>
</tr>
<tr>
<td>Belgium</td>
<td>1570</td>
<td>-1</td>
<td>1671</td>
<td>655</td>
<td>625</td>
<td>877</td>
</tr>
<tr>
<td>Denmark</td>
<td>400</td>
<td>407</td>
<td>496</td>
<td>-</td>
<td>265</td>
<td>679</td>
</tr>
<tr>
<td>Finland</td>
<td>1119</td>
<td>-1</td>
<td>435</td>
<td>112</td>
<td>253</td>
<td>232</td>
</tr>
<tr>
<td>France</td>
<td>25520</td>
<td>17293</td>
<td>14087</td>
<td>3989</td>
<td>2611</td>
<td>1528</td>
</tr>
<tr>
<td>Germany</td>
<td>26176</td>
<td>17381</td>
<td>8411</td>
<td>6245</td>
<td>3849</td>
<td>14592</td>
</tr>
<tr>
<td>Greece</td>
<td>2802</td>
<td>5192</td>
<td>978</td>
<td>266</td>
<td>320</td>
<td>3355</td>
</tr>
<tr>
<td>Hungary</td>
<td>6075</td>
<td>792</td>
<td>1163</td>
<td>180</td>
<td>191</td>
<td>-1</td>
</tr>
<tr>
<td>Ireland</td>
<td>263</td>
<td>401</td>
<td>312</td>
<td>352</td>
<td>165</td>
<td>240</td>
</tr>
<tr>
<td>Italy</td>
<td>13645</td>
<td>-1</td>
<td>7229</td>
<td>2784</td>
<td>2324</td>
<td>4145</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>105</td>
<td>-1</td>
<td>61</td>
<td>33</td>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>Netherlands</td>
<td>4411</td>
<td>3103</td>
<td>1482</td>
<td>311</td>
<td>686</td>
<td>1421</td>
</tr>
<tr>
<td>Portugal</td>
<td>1791</td>
<td>121</td>
<td>472</td>
<td>212</td>
<td>483</td>
<td>501</td>
</tr>
<tr>
<td>Spain</td>
<td>6224</td>
<td>3312</td>
<td>2067</td>
<td>2965</td>
<td>1474</td>
<td>2307</td>
</tr>
<tr>
<td>Sweden</td>
<td>2172</td>
<td>-1</td>
<td>456</td>
<td>143</td>
<td>383</td>
<td>953</td>
</tr>
<tr>
<td>Switzerland</td>
<td>4030</td>
<td>587</td>
<td>532</td>
<td>521</td>
<td>202</td>
<td>925</td>
</tr>
<tr>
<td>UK</td>
<td>12728</td>
<td>19371</td>
<td>5192</td>
<td>5768</td>
<td>2392</td>
<td>1994</td>
</tr>
<tr>
<td>Total</td>
<td>113413</td>
<td>69515</td>
<td>45877</td>
<td>24865</td>
<td>16295</td>
<td>35172</td>
</tr>
</tbody>
</table>

1 No data
2 Lisbon and Oportom metropolitan areas only
Table 2  
Road revenues and taxes  
- € million 1998- 

<table>
<thead>
<tr>
<th></th>
<th>Charges for infrastructure use</th>
<th>Vehicle taxes</th>
<th>Fuel taxes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed</td>
<td>Variable</td>
<td>Registration tax</td>
<td>Circulation tax</td>
</tr>
<tr>
<td>Austria</td>
<td>266</td>
<td>237</td>
<td>834</td>
<td>391</td>
</tr>
<tr>
<td>Belgium</td>
<td>95</td>
<td>18</td>
<td>284</td>
<td>1153</td>
</tr>
<tr>
<td>Denmark</td>
<td>131</td>
<td>38</td>
<td>2439</td>
<td>725</td>
</tr>
<tr>
<td>Finland</td>
<td>0</td>
<td>0</td>
<td>1262</td>
<td>184</td>
</tr>
<tr>
<td>France</td>
<td>0</td>
<td>4167</td>
<td></td>
<td>4983</td>
</tr>
<tr>
<td>Germany</td>
<td>411</td>
<td>0</td>
<td>7757</td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td>1327</td>
<td>0</td>
<td>280</td>
<td>741</td>
</tr>
<tr>
<td>Hungary</td>
<td>122</td>
<td>12</td>
<td>31</td>
<td>76</td>
</tr>
<tr>
<td>Ireland</td>
<td>0</td>
<td>27</td>
<td>770</td>
<td>373</td>
</tr>
<tr>
<td>Italy</td>
<td>2222</td>
<td>865</td>
<td>3325</td>
<td>934</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>3</td>
<td>1</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>Netherlands</td>
<td>91</td>
<td>0</td>
<td>1873</td>
<td>2425</td>
</tr>
<tr>
<td>Portugal</td>
<td>52</td>
<td>332</td>
<td>1030</td>
<td>63</td>
</tr>
<tr>
<td>Spain</td>
<td>0</td>
<td>0</td>
<td>1266</td>
<td>139</td>
</tr>
<tr>
<td>Sweden</td>
<td>59</td>
<td>0</td>
<td>684</td>
<td>30</td>
</tr>
<tr>
<td>Switzerland</td>
<td>266</td>
<td>0</td>
<td>1041</td>
<td>125</td>
</tr>
<tr>
<td>UK</td>
<td>259</td>
<td>0</td>
<td>7500</td>
<td>132</td>
</tr>
</tbody>
</table>

1) Vehicle import tax. - 2) Also includes VAT on import tax and circulation tax. - 3) Sales tax. - 4) Also includes VAT on infrastructure charges. - 5) Insurance tax. - 6) Not included in this total are subsidies payments received by private motorway concessionaires for exchange rate risk totalling €197 million in 1998. - 7) All vehicle taxes: registration tax, insurance tax, taxes on company cars, tax on the vignette and tax on vehicle parts. - 8) Bus fuel duty rebate of €398 million has been deducted from this total. - 9) Insurance and radio tax. - 10) All other vehicle taxes. - 11) Municipal vehicle tax. - 12) Not included are subsidies granted for the provision of infrastructure totalling €171 million in 1998. - 13) None reported within the country account.
Table 3. Total rail transport costs, € million 1998

<table>
<thead>
<tr>
<th>Country</th>
<th>Infrastructure costs</th>
<th>Air pollution costs</th>
<th>Noise costs</th>
<th>Costs of global warming</th>
<th>External costs of accidents</th>
<th>Transport operator costs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>1933</td>
<td>15</td>
<td>6</td>
<td>7</td>
<td>23</td>
<td>2183</td>
<td>4 167</td>
</tr>
<tr>
<td>Belgium</td>
<td>1142</td>
<td>19</td>
<td>38</td>
<td>11</td>
<td>2</td>
<td>2579</td>
<td>3 791</td>
</tr>
<tr>
<td>Denmark</td>
<td>255</td>
<td>12</td>
<td>2)</td>
<td>9</td>
<td>21</td>
<td>795</td>
<td>1 092</td>
</tr>
<tr>
<td>Finland</td>
<td>360</td>
<td>7</td>
<td>22</td>
<td>6</td>
<td>5</td>
<td>451</td>
<td>851</td>
</tr>
<tr>
<td>France</td>
<td>4790</td>
<td>62</td>
<td>51</td>
<td>16</td>
<td>3</td>
<td>10944</td>
<td>15 916</td>
</tr>
<tr>
<td>Germany</td>
<td>12621</td>
<td>220</td>
<td>1031</td>
<td>152</td>
<td>83</td>
<td>7336</td>
<td>21 443</td>
</tr>
<tr>
<td>Greece</td>
<td>390</td>
<td>6</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>326</td>
<td>736</td>
</tr>
<tr>
<td>Hungary</td>
<td>505</td>
<td>41</td>
<td>27</td>
<td>6</td>
<td>2)</td>
<td>432</td>
<td>1 011</td>
</tr>
<tr>
<td>Ireland</td>
<td>221)</td>
<td>8</td>
<td>29</td>
<td>2)</td>
<td>2</td>
<td>255</td>
<td>316</td>
</tr>
<tr>
<td>Italy</td>
<td>5605</td>
<td>145</td>
<td>243</td>
<td>61</td>
<td>10</td>
<td>6673</td>
<td>12 737</td>
</tr>
<tr>
<td>Luxembourg3)</td>
<td>90</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2)</td>
<td>294</td>
<td>389</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1095</td>
<td>10</td>
<td>22</td>
<td>2</td>
<td>58</td>
<td>2339</td>
<td>3 526</td>
</tr>
<tr>
<td>Portugal</td>
<td>292</td>
<td>22</td>
<td>5</td>
<td>3</td>
<td>11</td>
<td>558</td>
<td>891</td>
</tr>
<tr>
<td>Spain</td>
<td>3500</td>
<td>50</td>
<td>219</td>
<td>27</td>
<td>19</td>
<td>2013</td>
<td>5 828</td>
</tr>
<tr>
<td>Sweden</td>
<td>856</td>
<td>5</td>
<td>43</td>
<td>3</td>
<td>32</td>
<td>1270</td>
<td>2 209</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2762</td>
<td>5</td>
<td>60</td>
<td>0.1</td>
<td>8</td>
<td>2095</td>
<td>4 930</td>
</tr>
<tr>
<td>UK</td>
<td>3288</td>
<td>343</td>
<td>107</td>
<td>54</td>
<td>26</td>
<td>6664</td>
<td>10 482</td>
</tr>
</tbody>
</table>

1) Operating, signalling and depreciation costs only. - 2) No data available for the estimation of these costs. 3) Rail owned buses included.
### Table 4. Rail revenues and subsidies. € million 1998

<table>
<thead>
<tr>
<th></th>
<th>Revenues</th>
<th>Taxes</th>
<th>Explicit subsidies</th>
<th>Implicit subsidies</th>
<th>Total (^7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ticket and freight revenues</td>
<td>Track, station and other infrastructure charges</td>
<td>Fuel and energy tax</td>
<td>For the provision of services</td>
<td>For concessionary fares</td>
</tr>
<tr>
<td>Austria</td>
<td>1277</td>
<td>349</td>
<td>5(^7)</td>
<td>1045</td>
<td>619</td>
</tr>
<tr>
<td>Belgium</td>
<td>908</td>
<td>5)</td>
<td>0.85</td>
<td>1615</td>
<td>5)</td>
</tr>
<tr>
<td>Denmark</td>
<td>566</td>
<td>20</td>
<td>0</td>
<td>219</td>
<td>30</td>
</tr>
<tr>
<td>Finland</td>
<td>533</td>
<td>54</td>
<td>4.8(^1)</td>
<td>53</td>
<td>9</td>
</tr>
<tr>
<td>France</td>
<td>6380</td>
<td>946</td>
<td>41</td>
<td>5678</td>
<td>296</td>
</tr>
<tr>
<td>Germany</td>
<td>8614</td>
<td>4566</td>
<td>251(^1)</td>
<td>7175</td>
<td>4244</td>
</tr>
<tr>
<td>Greece</td>
<td>126</td>
<td>5)</td>
<td>9(^1)</td>
<td>5)</td>
<td>126</td>
</tr>
<tr>
<td>Hungary</td>
<td>84</td>
<td>124</td>
<td>27</td>
<td>295</td>
<td>5)</td>
</tr>
<tr>
<td>Ireland</td>
<td>127</td>
<td>5)</td>
<td>5)</td>
<td>42</td>
<td>4)</td>
</tr>
<tr>
<td>Italy</td>
<td>3441(^3)</td>
<td>5)</td>
<td>5)</td>
<td>1740</td>
<td>1700</td>
</tr>
<tr>
<td>Luxembourg(^5)</td>
<td>100</td>
<td>5)</td>
<td>0.4</td>
<td>104</td>
<td>5)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1210</td>
<td>155</td>
<td>5)</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>Portugal</td>
<td>188</td>
<td>5)</td>
<td>5)</td>
<td>10</td>
<td>5)</td>
</tr>
<tr>
<td>Spain</td>
<td>1495</td>
<td>5)</td>
<td>0</td>
<td>1925</td>
<td>2(^2)</td>
</tr>
<tr>
<td>Sweden</td>
<td>1325</td>
<td>98</td>
<td>5)</td>
<td>500</td>
<td>5)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2191</td>
<td>774</td>
<td>5)</td>
<td>1621</td>
<td>5)</td>
</tr>
<tr>
<td>UK</td>
<td>5677</td>
<td>3448</td>
<td>5)</td>
<td>43</td>
<td>2254</td>
</tr>
</tbody>
</table>

\(^{1}\) Including VAT on fuel tax. \(^{2}\) Unknown level of subsidies for concessionary fares included in subsidies for provision of services. \(^{3}\) Including revenues of €1517 million from public service contract, which may also be seen as a subsidy. \(^{4}\) Revenues, taxes and subsidies from rail owned buses included. \(^{5}\) None recorded within the country account. \(^{6}\) Can not be calculated with the available data. \(^{7}\) Excluding infrastructure charges and implicit subsidies.
Table 5. Total air transport costs. € million 1998

<table>
<thead>
<tr>
<th>Country</th>
<th>Infrastructure Costs</th>
<th>Air pollution costs</th>
<th>Noise costs</th>
<th>Costs of global warming</th>
<th>External costs of accidents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>509</td>
<td>29</td>
<td>3</td>
<td>41</td>
<td>6</td>
<td>588</td>
</tr>
<tr>
<td>Belgium</td>
<td>184</td>
<td>11</td>
<td>:</td>
<td>116</td>
<td>0,85</td>
<td>312</td>
</tr>
<tr>
<td>Denmark</td>
<td>293</td>
<td>7</td>
<td>:</td>
<td>9</td>
<td>2</td>
<td>329</td>
</tr>
<tr>
<td>Finland</td>
<td>:</td>
<td>4</td>
<td>:</td>
<td>17</td>
<td>0,2</td>
<td>21</td>
</tr>
<tr>
<td>France</td>
<td>8110</td>
<td>60</td>
<td>:</td>
<td>31</td>
<td>0</td>
<td>8201</td>
</tr>
<tr>
<td>Germany</td>
<td>3488</td>
<td>162</td>
<td>278</td>
<td>434</td>
<td>35</td>
<td>4397</td>
</tr>
<tr>
<td>Greece</td>
<td>239</td>
<td>6</td>
<td>24</td>
<td>0,03</td>
<td>:</td>
<td>269</td>
</tr>
<tr>
<td>Hungary</td>
<td>127</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>141</td>
</tr>
<tr>
<td>Ireland</td>
<td>401</td>
<td>20</td>
<td>:</td>
<td>57</td>
<td>:</td>
<td>478</td>
</tr>
<tr>
<td>Italy</td>
<td>571</td>
<td>77</td>
<td>193</td>
<td>197</td>
<td>2</td>
<td>1041</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>37</td>
<td>1</td>
<td>:</td>
<td>2</td>
<td>:</td>
<td>40</td>
</tr>
<tr>
<td>Netherlands</td>
<td>98</td>
<td>25</td>
<td>186</td>
<td>15</td>
<td>0,4</td>
<td>325</td>
</tr>
<tr>
<td>Portugal</td>
<td>203</td>
<td>106</td>
<td>4</td>
<td>50</td>
<td>1</td>
<td>363</td>
</tr>
<tr>
<td>Spain</td>
<td>411</td>
<td>62</td>
<td>188</td>
<td>208</td>
<td>4</td>
<td>873</td>
</tr>
<tr>
<td>Sweden</td>
<td>447</td>
<td>2</td>
<td>0,4</td>
<td>65</td>
<td>1</td>
<td>515</td>
</tr>
<tr>
<td>Switzerland</td>
<td>804¹⁷</td>
<td>17</td>
<td>27</td>
<td>34</td>
<td>10</td>
<td>738</td>
</tr>
<tr>
<td>UK</td>
<td>2236</td>
<td>656</td>
<td>155</td>
<td>49</td>
<td>5</td>
<td>3101</td>
</tr>
</tbody>
</table>

¹⁷Including the costs of air traffic management services totalling €154 million in 1998.

Source: Link et al. (2002a,b,c)
Table 6. Revenues, charges, taxes and subsidies within the aviation sector (1998). € million

<table>
<thead>
<tr>
<th>Country</th>
<th>Airport revenues</th>
<th>ATM charges</th>
<th>Revenues lost: VAT on ticket price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>278</td>
<td>151</td>
<td>25&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>Belgium</td>
<td>255</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>:</td>
<td>:</td>
<td>103</td>
</tr>
<tr>
<td>Finland</td>
<td>181</td>
<td>:</td>
<td>0.3</td>
</tr>
<tr>
<td>France</td>
<td>1687</td>
<td>1117</td>
<td>279&lt;sup&gt;7&lt;/sup&gt;</td>
</tr>
<tr>
<td>Germany</td>
<td>3121</td>
<td>767&lt;sup&gt;1&lt;/sup&gt;</td>
<td>48&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Greece</td>
<td>767&lt;sup&gt;7&lt;/sup&gt;</td>
<td>:</td>
<td>34</td>
</tr>
<tr>
<td>Hungary</td>
<td>103</td>
<td>:</td>
<td>2</td>
</tr>
<tr>
<td>Ireland</td>
<td>134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>795</td>
<td>200</td>
<td>12</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>11</td>
<td>1.1</td>
<td>0</td>
</tr>
<tr>
<td>Netherlands</td>
<td>224</td>
<td></td>
<td>1.3</td>
</tr>
<tr>
<td>Portugal</td>
<td>114</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>501</td>
<td>341</td>
<td>77&lt;sup&gt;6&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sweden</td>
<td>184</td>
<td>119</td>
<td>17</td>
</tr>
<tr>
<td>Switzerland</td>
<td>651</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>:</td>
<td>137&lt;sup&gt;3&lt;/sup&gt;</td>
<td>1210&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>1</sup>Meteorological services charge. <sup>2</sup>For Lufthansa only. <sup>3</sup>Security charge. <sup>4</sup>Air passenger duty. <sup>5</sup>All airport and ATM charges. <sup>6</sup>Subsidies to airlines. <sup>7</sup>€194 million to airports, € 85 million other general subsidies. <sup>8</sup>Profit from these services going to general budget.

Source: Link et al. (2002a,b,c)
References to Appendix A


Appendix B. Congestion and Scarcity costs

*Working paper for D2 of CATRIN*

*Pedro Abrantes and Chris Nash (ITS Leeds)*

**Introduction**

The current structure of rail track access charges in Europe takes little account of the congestion and infrastructure scarcity costs of running train services. Although these costs are often negligible they can become quite significant where and when networks are operating close to capacity. The achievement of allocative efficiency requires track access charges to be based on Marginal Social Cost (MSC) in line with EC Directive 2001/14. Ignoring congestion and scarcity costs can introduce distortions in the market, for example by constraining the reallocation of train paths to competing uses or encouraging the over-supply of services. It also leads to much lower cost recovery levels than would otherwise be achieved, with the attendant costs in terms of the cost of public funds, resulting from distortions introduced by the tax system.

<table>
<thead>
<tr>
<th>Country</th>
<th>Type of charge</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>Reservation charge per path-km (up to 15€/km) and station stop (up to 26€/stop), differentiated by line type and time period</td>
<td>Scarcity</td>
</tr>
<tr>
<td>Hungary</td>
<td>Charge per train-path (3 passenger categories and 1 freight)</td>
<td>Scarcity</td>
</tr>
<tr>
<td>Italy</td>
<td>Fixed reservation charge by line type plus variable charge dependent on traffic density</td>
<td>Scarcity/ Congestion</td>
</tr>
<tr>
<td>Lithuania</td>
<td>Train path reservation fee for freight</td>
<td>Scarcity</td>
</tr>
<tr>
<td>Denmark</td>
<td>Higher charges on special infrastructure points (bottlenecks)</td>
<td>Scarcity/ Congestion/ Cost recovery?</td>
</tr>
<tr>
<td>UK, Romania, Germany, Austria</td>
<td>Congestion taken into account through charge differentiation</td>
<td>Congestion</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>Fixed path reservation fee</td>
<td>Path planning</td>
</tr>
<tr>
<td>Estonia</td>
<td>Fixed charge per ordered train-km</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 highlights some examples of current charges possibly linked to congestion and scarcity costs. The UK has an explicit capacity charge (Gibson et al, 2002) whereas Austria, Germany, Denmark and Italy all have some kind of additional bottleneck charge applied at certain sections of the network where capacity may be constrained. France’s highly differentiated train path reservation fees are particularly interesting (Table 2) and may implicitly recognize the opportunity cost of a given train-path, given that the differentiation by line type takes into account traffic density, and the charging of different fees according to peak, normal and off peak times of day.
**Table 2. Breakdown of Track Access and Reservation Charges for 2009 (France), Source: RFF, Network Management Statement, 2007**

<table>
<thead>
<tr>
<th>Category</th>
<th>A (Urban and suburban lines, high traffic)</th>
<th>B (Urban and suburban lines, average traffic)</th>
<th>C &amp; C* (Main interurban lines, high traffic (* = 220 km pH))</th>
<th>D &amp; D* (Main interurban lines, average traffic)</th>
<th>E (Other lines)</th>
<th>N1 (High speed lines, high level traffic)</th>
<th>N2 &amp; N2* (High speed lines, average traffic)</th>
<th>N3 &amp; N3* (High speed lines, low level traffic)</th>
<th>N4 (High speed lines, HSL Mediterranéen)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>287 km</td>
<td>985 km</td>
<td>7,209 km</td>
<td>5,840 km</td>
<td>12,738 km</td>
<td>718 km</td>
<td>12,42 km</td>
<td>5,554 km</td>
<td>321 km</td>
</tr>
<tr>
<td>Length (km)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Length = 457 km</td>
<td>Length = 457 km</td>
<td>Length = 321 km</td>
<td>Length = 300 km</td>
</tr>
<tr>
<td>Length (km)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Length = 718 km</td>
<td>Length = 12,42 km</td>
<td>Length = 5,554 km</td>
<td>Length = 321 km</td>
</tr>
<tr>
<td>Length (km)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Length = 718 km</td>
<td>Length = 457 km</td>
<td>Length = 321 km</td>
<td>Length = 300 km</td>
</tr>
</tbody>
</table>

**Fixed track access charges (monthly fee per path km)**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.015</td>
<td>0.015</td>
<td>0.015</td>
<td>0.0</td>
<td>0.0</td>
<td>1.051</td>
<td>1.051</td>
</tr>
</tbody>
</table>

**Track reservation charges**

<table>
<thead>
<tr>
<th></th>
<th>Off hours</th>
<th>Peak Non-Hours</th>
<th>Off hours</th>
<th>Peak Non-Hours</th>
<th>Off hours</th>
<th>Peak Non-Hours</th>
<th>Off hours</th>
<th>Peak Non-Hours</th>
<th>Off hours</th>
<th>Peak Non-Hours</th>
<th>Off hours</th>
<th>Peak Non-Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.89</td>
<td>0.76</td>
<td>0.765</td>
<td>0.0</td>
<td>0.0</td>
<td>6.1</td>
<td>1.427</td>
<td>1.024</td>
<td>0.908</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.14</td>
<td>1.479</td>
<td>0.765</td>
<td>0.052</td>
<td>0.005</td>
<td>12.42</td>
<td>5.554</td>
<td>2.842</td>
<td>2.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.9</td>
<td>3.35</td>
<td>1.58</td>
<td>0.052</td>
<td>0.005</td>
<td>15.14</td>
<td>7.95</td>
<td>4.792</td>
<td>4.376</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.07</td>
<td>5.61</td>
<td>5.61</td>
<td>5.61</td>
<td>5.61</td>
<td>9.07</td>
<td>5.61</td>
<td>5.61</td>
<td>5.61</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(*) 0H30-4H30
(**) 6H30-9H and 17H-21H

It must be said, however, that pricing according to what the market will bear might well follow a similar sort of structure, and the French charges also appear to be linked to ability to pay (there are thus high charges on some high speed lines even where capacity is not scarce and where there is only one operator permitted). The use of reservation charges (where operators are charged a certain amount for a slot for every occasion for which it is reserved for them, regardless of whether they end up using that train-path or not) is only appropriate in terms of reflecting costs in the case of scarcity costs, where other train operators are deprived of the use of the slot whether the operator booking it uses it or not. Most other aspects of marginal social costs are only incurred if the slot is actually used, whilst train planning costs are
more related to the number of time a path is planned rather than the number of occasions on which it is reserved.

**Congestion versus Scarcity**

Congestion costs are incurred when running an additional service on the network causes additional delays to other services even when that additional service does not itself directly cause any delay (where it is the direct cause of delay, for instance due to mechanical problems, the costs should be covered by the performance regime). When the infrastructure is being run close to capacity unexpected delays to one service are likely to lead to additional delays to other trains as there is little slack in the timetable to allow the system to recover. Gibson et al (2002) estimated marginal congestion costs in Britain by regressing delays against the level of capacity utilization. Their results suggested congestion charges should be as high as 5€ per train-km outside London and even higher for routes into London where capacity is at a premium. The congestion charge currently in place is much simpler than that originally proposed and has been criticized for not accurately reflecting marginal cost. However, the original work served to highlight that congestion costs can be a significant proportion of the marginal cost of train operations.

Scarcity costs arise where the operation of a train service prevents another from operating, or requires it to take an inferior path. Because a network might be used to run many possible combinations of services, it is difficult to uniquely determine its capacity and therefore the opportunity cost of a given service. Furthermore, the impact of an additional train of a particular type on the paths available to other trains can differ significantly according to the precise mix of traffic on the line. At the same time, the value of a standard slot to a given operator will also vary in time and space, and will depend on what else is running on the network. Quinet (2003) illustrates how the allocation of train paths to competing versus complementary services can have very different values. Given all these arguments the development of a general methodology for estimating the marginal scarcity cost of a train-path represents a significant research challenge.

It should be noted that both congestion and scarcity charges are already internalized when the operator of the train delayed or prevented from using the system is the same as the one causing the delay or lack of capacity. This leads to a complication in that, for instance, in a system with a single dominant operator but a number of smaller ones, the optimal charge per slot for the dominant operator is likely to be much less than for the small operator. Whether such a charging structure is permissible in terms of being non discriminatory is doubtful.

Nilsson (2002) suggests that the potential conflicts of interest between operators and the difficulties in establishing optimal track allocation procedures may have been one of the reasons for early railway entrepreneurs to eventually opt for a vertically and horizontally integrated industry structure. In any case, the re-structuring of the railway industry has generated some recent research on scarcity costs. Three main approaches have been proposed for establishing the scarcity value of a rail slot (Nash et al, 2005):

- Auctioning
- Use of models to determine the opportunity cost of an additional service
- Use of the long term average incremental cost of capacity expansion as a proxy (NERA, 1998)

**Auctioning**
In principle, the most attractive method of estimating the opportunity cost of a slot in terms of what another operator is willing to pay for it would be to auction available slots. There are many practical difficulties however, including picking the optimal allocation out of the multiple combinations of slots possible and the fact that the value of a particular slot for a particular purpose depends on how other slots are being used.

There has been some research in the field of experimental economics into the applicability of theoretical auctioning procedures to the allocation of rail slots. The seminal work by Brewer and Plott (1996) involved the design and experimental test of a first-price bidding mechanism for conflicting train-paths within a simple stylized rail network. The binary conflict ascending price (BICAP) proposed is essentially a traditional English auction whereby after successive bidding rounds the operator with the highest bid is allocated the slot at a price equal to the winning bid. The experiments carried out by the authors showed that this type of auction can achieve a slot allocation close to the theoretical optimum.

Nevertheless, Nilsson (1999) argues that a second-price mechanism (Vickrey, 1961) will produce a stronger incentive for operators to bid closer to their real valuation and therefore lead to more efficient allocations and higher selling prices. In this type of auction the highest bidder pays a price equivalent to the maximum of the sum of the bids of the combination of services that could otherwise have been run on the system. Results also show this mechanism to be able to produce highly efficient allocations under experimental conditions.

Isaacsson and Nilsson (2003) compared first, second-price and one-shot auctions in an additional set of experiments. Although the latter is likely to be more effective against collusion it also tended to produce less efficient allocations than the alternative procedures. Both first and second-price auctions produced relatively efficient allocations even though the second-price mechanism provides more efficient solutions where the conflicts generated gave rise to free riding incentives.

In contrast with the above-mentioned studies, experiments by Cox et al (2002) showed a first-price auction to produce significantly lower consumer and producer surplus than a central timetable scheduling procedure making use of operators’ true valuations. The reason behind this discrepancy could be the greater complexity of the test network. It is easy to see that as the complexity of the network increases a central scheduling procedure will become noticeably more effective at determining the optimum slot allocation than a non-cooperative auctioning procedure. On the other hand, the problem remains of how good the information of the central planner is about the true valuations. One problem in comparing these lab-based experiments is that conditions have tended to be very different.

Another criticism that can be levied at experimental auction tests is that they necessarily rely on very simplified networks and subjects without specialist knowledge of the rail industry. It remains to be seen what the outcome would be for an experiment where rail professionals were faced with a realistic network allocation problem.

Quinet (2003) points out that even if such a large-scale auction was deemed to produce an efficient allocation of resources the transaction costs involved in reaching an equilibrium solution would make this impractical in any but the simplest of networks. The author therefore proposes that one-shot auctions be combined with central planning to produce final allocations.
Nilsson (2002) proposes an iterative procedure whereby operators bid for their preferred set of paths and also state how much less they would be willing to pay for an inferior solution. The infrastructure manager would then use optimization procedures to determine the best timetable to maximize producer welfare. Having received their allocation of slots the operators would have a chance to modify their bids and the process would continue until equilibrium was reached. This procedure allows for the fact that the willingness of an operator to pay for a particular slot will depend on what other slots that operator gets and what slots their rivals get.

To the extent that this is not unlike the iterative process already adopted by infrastructure managers for resolving conflicts, it appears feasible, although the big unknown is how long the system would take to converge to equilibrium. Nilsson recognizes that the approach is not suitable where operators are competing for the same traffic (since their willingness to pay would depend on how much traffic they could divert from their rival, and this may be very different from the social benefit of their gaining the slot) but would be effective for dealing with conflicts between operators of different types of traffic.

It should be noted that a second price auction will automatically achieve the result that the opportunity cost of a slot is only charged where it results from the willingness to pay of another operator, as firms will not bid against themselves. A first price auction may lead to a price above the opportunity cost of the slot, although obviously it will always be equal to or less than the willingness to pay of the firm.

**Modelling**

Another approach to measuring the opportunity cost of a given train-path is to use predictive models and *a priori* knowledge of passengers’ willingness to pay to forecast the impact on demand, revenue and social welfare of a given slot allocation rather than to obtain that valuation from operators through auctioning. The opportunity cost of a given slot can be estimated as the sum of:

- the additional amount of traffic attracted to rail by the presence of a given train multiplied by the price it pays
- the consumers surplus to rail users as a result of the additional quality and capacity provided by the train
- the savings of external costs to road users and the public at large from the train attracting passengers from road
- less the train operating, infrastructure and external cost savings from failing to run this train

Nash et al. (2004) illustrate this approach with a worked example for the British Transpennine Route, and Johnson and Nash (2005) use the PRAISE model (Whelan et al., 1997, Preston et al., 1999) to estimate the value of peak slots on the British East Coast Mainline using this approach. PRAISE is a very detailed model able to estimate demand for individual services and ticket types, as a function of fares, journey times, desired departure times and overcrowding. Provided data is available and the model is accurately calibrated this offers an attractive approach to determining scarcity costs based on producer and consumer surplus as well as external costs and benefits. Indeed, the results of this work highlighted that auctioning may not give the optimum result without the payment of subsidies to operators due to the significant proportion of external benefits which different path allocations can generate.
**Long term average incremental cost of capacity expansion**

Another approach, recommended by NERA (1998), is to identify sections of infrastructure where capacity is constrained and to charge the long run average incremental cost of expanding capacity. However, this is a very difficult concept to measure (the cost of expanding capacity varies enormously according to the exact proposal considered, and it is not easy to relate this to the number of paths created, since they depend on the precise number and order of trains run). It may be argued, however, that more appropriate incentives are given to infrastructure managers if they are allowed to charge the costs of investment they actually undertake, rather than for the scarcity resulting from a lack of investment, at least if they are commercially oriented.

**Conclusions**

The implementation of scarcity charges makes sense where there are competing uses for rail infrastructure. An interesting application of this concept would be in determining an adequate charge for possessions regimes, where there is a clear opportunity cost involved.

However, there are significant problems in the estimation of scarcity costs for competing services. Given the difficulties with all the approaches discussed, an attractive solution may be to permit direct negotiation between operators and the infrastructure manager over the price and allocation of slots, including investment in new or upgraded capacity. British experience of this approach is that it is complex, time consuming and will not necessarily lead to an optimal outcome given the number of parties involved and the scope for free-riding. In countries where the infrastructure manager and incumbent operator have an affinity there may also be questions regarding transparency and the opportunity for discriminatory behaviour.

The French approach, which includes reservation charges based on differentiation of charges by time period, and traffic density on the route in question offers a pragmatic solution although there is no evidence that these charges accurately reflect scarcity costs. Nevertheless even a crude approximation may offer benefits in terms of giving some incentive to avoid the most heavily booked times and places, as well as raising the level of cost recovery. Scarcity charges should be distinguished from mark ups designed purely to improve cost recovery rates as the former actually improve incentives for efficient use of the infrastructure whereas the latter distort them.  

In terms of congestion, cost estimation can be relatively straightforward as illustrated by Gibson et al (2002). However, even in the UK case there have been practical barriers to implementing the highly disaggregate congestion charges proposed by the authors. For instance, the necessary information for billing must be collected, whilst freight operators complain that they do not know what they will be charged until they know what slot they have been allocated, by which time they may well have agreed a price with their customers. Thus so far, only a simpler charge, not varying by time of day, has been implemented, although it is intended to move to a more sophisticated system following the current periodic review of charges.

Finally, it should be reiterated that, where congestion or scarcity costs are imposed on the same firm as causes the cost, they are already internalised. Thus they are only really important in pricing when there is a lot of competition between different firms (or between parts of the same firm that act independently, such as passenger and freight sectors of a traditional railway company).
References


Appendix C Cost recovery for transport infrastructure

CATRIN Task 2.3

Jeremy Toner and Chris Nash, Institute for Transport Studies, Leeds

This note covers three possibilities for departures from marginal cost pricing to permit cost recovery in decreasing-cost situations. We ignore externalities; the extension of the concept of marginal cost to include such is essentially trivial, although if the revenue from such charges is available to the infrastructure manager then they do affect the strength of the budget constraint. Likewise, without loss of generality, we use ‘price’ in the conventional economist’s way, acknowledging that the concept is readily extendable to ‘full price’ or generalised cost.

Cost recovery has been an issue since before Dupuit, although his treatment can be considered the first treatment of the question of economically efficient solutions to maximising social welfare subject to a cost recovery constraint (Dupuit, 1849). Other texts of the time (for example Lardner, 1850) also seem to include cost recovery as a desirable principle; the distinction is that such efforts rely on notions of cost allocation. The neo-classical economist would say that concepts such as ‘fully distributed’ cost (accounting cost) or ‘full cost of the lowest cost provider’ should not be entertained since they depart from marginalist principles.

Introduction to concepts

In this section, we briefly review the three major approaches to cost recovery: fully distributed cost pricing; Ramsey pricing; and multi-part tariffs.

Fully Distributed Cost pricing

One approach has been to specify Fully Distributed Cost (FDC) prices, so that each output generates at least enough revenue to cover the costs of producing it. The difficulty comes with shared or common costs which cannot be uniquely attributed to one output. This is the basic problem with financing roads using the old British "Road Track Costs" mechanism, which allocated some 85% of the capital costs of road construction in proportion to the amount of roadspace each vehicle takes up. But any method for distributing these costs between outputs is essentially arbitrary; no one cost-allocation method is demonstrably theoretically superior to any other. FDC methods also eliminate any notion of marginality and they are inefficient, in that prices can be changed and welfare improved to meet a given budget constraint. On those grounds, we take this method no further.

Ramsey Pricing

Ramsey pricing can be used to cross-subsidise from areas of stronger demand to areas of more elastic demand (assuming the latter do, in general, at least cover short run marginal costs; although we note that in the case where the Ramsey-priced good is only part of a process, and there exist down-stream or up-stream price distortions, then price below marginal cost can feature). Note that in situations where, because of indivisibility, some parts of a system are capacity constrained, whereas others have excess capacity, it is an empirical matter as to
whether the surpluses (above long run marginal cost) from areas operating at the lower quantity dictated by the indivisibility are sufficient to cover the deficits (below long run marginal cost) from areas operating at the higher quantity.

Multipart (second-degree discriminatory) pricing

Whereas Ramsey pricing seeks to charge different prices to different (sub)markets identified, inter alia, by different price elasticities of demand, multipart or nonuniform pricing seeks to charge different prices to consumers in the same (sub)market, such that the individual consumer’s total spend on the good in question does not rise proportionately with consumption. In its simplest form, with an entry fee plus a per unit charge, the average price per unit falls with consumption; but, given that the amount consumed is positive, it is the marginal price which affects consumption decisions. More generally, it is possible to devise multipart tariffs with quantity premia as well as quantity discounts.

Meeting Budget Constraints – Second Best (Ramsey) Pricing

Suppose we have a multi-product firm which is unable/unwilling to price at marginal cost and does not wish to implement multi-part tariffs. There is a range of pricing possibilities. The firm may seek to profit-maximise. In such circumstances, it will adopt third-degree price discriminatory pricing as outlined above. If the firm is a monopoly and able to extract substantial surpluses, there may be regulatory intervention to reduce the profits; the firm may be set a maximum profit or, in the case of a publicly-owned utility, may be required to make normal economic profit.

The technique of Ramsey pricing gives rules for optimal (in the sense of least-distorting) departures from marginal cost pricing. In this case, we have assumed independence in supply, that is, although average costs decrease over the range of output for each good, there are no economies of scope. Figure 1 illustrates, in the two product case, two sets of prices consistent with break-even: set \{A,B\} is pricing to break-even in each market; set \{C,D\} is pricing to break-even over all. The end financial result is the same; but in welfare terms, the social surplus is greater under \{C,D\} where the profits from market 1 are used to cover some of the costs in market 2. Note that market 2 does pay above marginal cost, so makes some contribution to meeting average cost. In this example, we could specify any feasible profit target and find multiple price combinations to meet that target. But not all price combinations have equal efficiency effect; Ramsey prices are guaranteed to minimise the loss of efficiency associated with departure from marginal cost pricing.
Figure 1: Ramsey pricing

We start with a simple case of independent demands. This might be a single infrastructure operator seeking to price various facilities but where there is no relationship between the demand for one facility and the demand for any other - and therefore no relationship between the demand for one 'final' transport product and another. Another case could be pricing a particular facility or piece of infrastructure to different operators which are operating in different markets (e.g. passenger versus freight) and where the facility in question is not congested.

The basic rule is to price up more in the relatively inelastic markets and to price up less in the relatively elastic markets. Where there are zero cross-elasticities between markets (independent demands) the rule in each market $i$ is:

$$\frac{P_i - MC_i}{P_i} = -\frac{\lambda}{1 + \lambda} \frac{1}{\mathcal{E}_i}$$

(1)

where $\lambda$ reflects the tightness of the budget constraint ($\lambda = 0$ at marginal cost pricing; $\lambda \to \infty$ at profit maximisation). In this case where all the cross-elasticities of demand are zero, the rule requires that price be set so that "its percentage deviation from marginal cost is inversely proportional to the item's price elasticity of demand" (Baumol and Bradford, 1970, p.267). Thus prices will be nearer to marginal costs for items which are relatively price elastic, and further from MC where demand is relatively price inelastic. Thus, insofar as markets are separable, operators should price up in relatively inelastic markets.
We may consider the above case to be approximately true at a fairly aggregate (say, national) level where the goods are independent in demand (intercity rail services, regional rail services, freight services) and where there is spare capacity on the network so that one type of traffic’s demand for access does not materially affect the supply available to other traffic types.

Strictly, of course, if one is talking about pricing access to infrastructure, then the demand for access is a derived demand, in the same way that passenger demand is generally a derived demand. While there may be a distribution of tastes and preferences among passengers making a particular trip, one can nevertheless reason that the transport operator’s elasticity of demand for access will be related to some aggregation of passengers’ demands for travel. Specifically, if an operator has no market power (or has already used it all by adopting profit-maximising price discrimination), then the operator’s willingness to pay for access will be equal to the aggregation of the passengers’ willingness to pay for access. It is unlikely that this is known. However, the ratio of elasticities approach can be used to deduce this elasticity. This simply states that, where a price for a product can be unbundled into a set of part-prices for part-products, then the ratio of any two part-elasticities equals the ratio of the prices.

Suppose the price of 100gm of ham is €2 and the price of a ham roll containing 100gm of ham is €3. If the price elasticity of demand for ham rolls is -0.6, then we can deduce that the price elasticity of demand for ham is -0.4.

So if we rework equation (1), denoting \( P_i \), \( MC_i \) and \( \varepsilon_i \) as the price, marginal cost and price elasticity of demand for access to infrastructure and letting \( R_i \) be the total cost of the service excluding the access charge and \( \eta_i \) be the elasticity of demand for the whole service, equation (1) becomes

\[
P_i = \frac{MC_i - k}{\frac{k}{1 + \eta_i}} R_i
\]

(2)

where \( k = \frac{\lambda}{(1+\lambda)} \) and \( \eta_i = (P_i + R_i)\varepsilon_i/P_i \). (A treatment similar to this is given in Morrison (1982) in the context of landing fees at uncongested airports.)

So as in the standard case, if we were comparing two facilities with independent demands, we would still conclude that price should be have a greater mark up above marginal cost for the facility with the relatively inelastic demand ceteris paribus. So where access fees were a constant proportion of total cost, the elasticity of demand for the final product would still drive things.

In the case where two independent final products have the same price elasticity of demand, the share which the access fee takes in total cost drives things: now, the lower the share of access fees in total cost, the higher the mark up all other things equal (because for a given final product elasticity, the deduced access fee elasticity is smaller when access fee is a lower part of total cost). So if, for example, the price elasticity of demand for train trips from London to Brussels were the same as the elasticity of demand between London and Marseille, then the mark up which Eurotunnel charges as trains pass through the channel tunnel should
be proportionately bigger for the Marseille train than for the Brussels train (assuming it costs more to run a train from London to Marseille than it does to run one from London to Brussels).

We turn now to the more complicated case on interdependent demands. We see two key areas where this might be applied. The first is where there is competing demand for access, such that one operator’s actions or decisions affect another’s traffic (e.g. an open access passenger operator entering against an incumbent). Here, the crucial determinant will be the degree of substitutability in consumption between the products offered by the rival operators. The second situation is where one regards the products of a single operator a consisting of links of interdependent demands. For example, the British East Coast mainline could be regarded as three sections: London to Doncaster, Doncaster to Newcastle and Newcastle to Edinburgh. Demand for a trip from London to Edinburgh can therefore be regarded as the demand for each of the links in turn. If the price of one link changes (e.g. London to Doncaster), that will affect demand for London to Edinburgh trips. Of course, in many cases, end-to-end trips to the final user (the passenger) are not priced in this manner (although strict mileage based charging does do this); but the intermediate good (access to the tracks between London and Doncaster) may have a different elasticity from – but co-dependent with – the intermediate good ‘access between Doncaster and Newcastle’. The situation is perhaps more clearly seen with private motorists. Consider someone who wishes to drive from London to Strasbourg. The demand for road space between London and the Channel Tunnel is clearly co-dependent with demand for the Channel Tunnel and the demand for rival ferry service from Dover, as well as with demand for Belgian (untolled) or French (tolved) motorways heading east.

Consider an operator who runs a network of interdependent services and has a remit to break even. Baumol and Bradford (1970), rediscovering the work of Frank Ramsey (1927 – these rules are often referred to as Ramsey Pricing), have shown that the optimal pricing rule for a product $i$ depends on the marginal cost of $i$, the cross-elasticities of demand for $i$ with respect to the prices of the other products, the own-price elasticity of demand for $i$, and the tightness of the budget constraint (the marginal opportunity cost of using public funds which could be spent elsewhere, to subsidise the operator). In general terms, the rule is

$$P_i = \frac{MC_i - \lambda \left( \sum_j \frac{\partial P_j}{\partial x_i} \cdot x_j \right)}{1 + \frac{\lambda}{1+\lambda \cdot \varepsilon_i}}$$

or

$$\frac{P_j - MC_j}{P_i} = \frac{-\lambda \cdot 1}{1+\lambda \cdot \varepsilon_i} - \frac{\lambda}{1+\lambda} \cdot \sum_{j \neq i} \frac{1}{\varepsilon_{ij}}$$

where $\varepsilon_i$ is the own price demand elasticity of $i$, $x_j$ is the amount of $j$, $\varepsilon_{ij}$ is the cross-price elasticity of demand for $j$ with respect to the price of $i$, and $\lambda$ is the net gain which would occur if the budget constraint were relaxed by one unit, i.e. the tightness of the budget constraint.
When the cross-elasticities are not zero, the rule is more difficult to apply, but the principle is that price exceeds marginal cost by a greater amount:

(a) the tighter the budget constraint;
(b) the lower the own-price elasticity;
(c) the higher the cross-elasticities (the greater the substitutability between \( i \) and the firm's other products).

It is, perhaps, slightly paradoxical that when departures from marginal cost pricing are least distorting because consumers have close substitutes to which they can switch almost costlessly, this is the precise case where it is hardest to effect and least able to raise extra revenues.

Equivalently, the rule is that "all outputs be reduced by the same proportion from the quantities which would be demanded at prices equal to the corresponding marginal costs" (op. cit., p.267). These results are for a multi-product operator, but apply equally to social welfare maximisation across the products of different suppliers (eg. car/bus/train).

To what extent does the intermediate nature of facility access alter this? To the extent that we can apply the ratio of elasticities approach, we can deduce price elasticities for the use of the facility. Of course, an infrastructure manager may have less scope for differentiating different markets than a transport undertaking. A transport undertaking may charge according to the origin and destination of the passenger and the passenger’s age, and indeed carry out individual negotiations with freight customers. At the most, an infrastructure manager’s scope for negotiation is at the level of the vehicle or train movement. Where such negotiation is infeasible or not allowed, then the infrastructure manager will be restricted to charging according to such factors as section of route, time of day and type of vehicle/train.

There is a further issue regarding scheduled transport services, in that frequency is an important attribute of service quality. Raising the charge per vehicle or train kilometre will therefore not simply raise price of the final good, but also the choice of frequency to offer, leading to additional welfare costs. A further potential difficulty comes if further market distortions upstream or downstream are (at least partially) to be corrected by facility pricing.

The effect of Ramsey pricing on social welfare depends on the tightness of the budget constraint. The further prices have to depart from marginal cost pricing to meet a constraint, the bigger the deleterious effect on social welfare. Nevertheless, Ramsey prices offer a least-distorting way of departing from marginal cost pricing to achieve a given constraint in the absence of multipart tariffs.

**Multipart tariffs – the use of second degree discrimination**

Unless the price elasticity of demand is zero in some markets, Ramsey pricing will always have a welfare cost. By contrast, a multipart tariff in which the fixed part of the tariff can be structured so that no customer leaves the market, and the variable part tracks marginal cost may increase revenue with no welfare cost at all. Arguably there may be circumstances in which the fixed access fee deters few if any users; for instance, almost all houses are connected to electricity supply. However in practice, the access fee may be distorting, which then leads to
the situation that the marginal unit depends on the level of access fee and so the exercise becomes one of joint determination of access fee and per-unit charge in order to maximise social welfare. For some consumers who are willing to pay marginal cost, the access fee may be greater than their consumer surplus, causing them to leave the market. It is, of course, theoretically possible to devise discriminating non-linear tariffs to allow for this.

Brown and Sibley (1986) present such a case. Let \( F \) be the firm’s deficit if it applied marginal cost pricing to its product. Then if there are \( n \) consumers, if a two-part tariff of \( E + pQ_i \) (where \( E = F/n; p \) is the marginal price and, in this instance is set equal to \( c \), the marginal cost; and \( Q_i \) units are consumed by the \( i \)th consumer) is charged to each consumer and if all consumers remain in the market, then an efficient two part tariff has been constructed and there is no change in total social surplus compared with standard marginal cost pricing; the only change is in distribution, with a reallocation of some consumer surplus to the firm.

One can see how this might also apply to intermediate goods, such as access to railway infrastructure. The precise detail of an efficient two part tariff in this case would depend on an assumption about the characteristics of the final good(s) market(s). If the final goods markets are perfectly competitive, then the Coase two part tariff outlined above remains optimal. We will, for simplicity, proceed on this basis.

Suppose there are two operators sharing a piece of infrastructure, for example suburban and long-distance operators sharing tracks and access for the last 100km into London. For as long as a fixed charge of \( E = F/2 \) keeps both operators in the market, the tariff is optimal. Suppose now that a new entrant wishes to offer just a small number of long distance services a day. If an entry charge of \( E = F/3 \) does not deter entry, then the Coase tariff remains optimal. It seems likely, though, that in many cases, if the entry charge is large relative to the proposed scale of entry, \( E \) would deter entry and this tariff would thus no longer be optimal.

One can use Ramsey principles to determine the relative balance between \( E \) and \( P \) according to their relative elasticities. If participation (or entry) is inelastic relative to unit consumption, then an optimal two part tariff will have a relatively high entry charge and a relatively low unit charge; and vice versa in the reverse case. But the setting of the participation price, \( E \), and the unit price, \( P \), must include consideration of the fact that the two goods are complements.

Two pricing rules (for participation and consumption) of a Ramsey form can be derived (see Brown and Sibley chapter 4). The results depend not only on prices, marginal costs and elasticities, but also on a distribution parameter. The crucial value of the distribution parameter is that of the marginal participant; for \( \theta \leq \theta_0 \), the individual does not participate (\( E \) deters participation/entry) and so consumes nothing and incurs no costs. For \( \theta \geq \theta_0 \), individuals participate. \( Q_0 = f(P, \theta_0) \), the amount consumed by the marginal consumer, is important, both in terms of allowing for the contribution of net consumption revenue from the marginal consumer in setting the access fee and through its role relative to average consumption (averaged only over those who consume) in setting the unit charge above or below marginal cost.

For entry/participation, the rule is

\[
\frac{\{E + PQ_0\} - \{v + cQ_0\}}{E} = \frac{M}{e}
\]

(5)
where $E$ is the participation price, $P$ is the unit consumption price and $c$ the unit consumption marginal cost, $v$ is the cost of allowing the marginal consumer to participate, $Q_0$ is the amount consumed by the marginal consumer, $M$ is a measure of the tightness of the budget constraint and $e$ is the price elasticity of demand of participation.

The similar rule for consumption price is

$$\frac{P - c}{P} = \frac{M}{\xi} \left( 1 - \frac{Q_0}{Q} \right)$$

where $Q'$ is the mean consumption of those who consume and $\xi$ is price elasticity of demand of total consumption (i.e. the elasticity of $Q$ with respect to $P$). Note that for $Q_0$ small relative to $Q'$, this is almost a standard Ramsey formula. But as $Q_0$ rises relative to $Q'$, the markup gets smaller and turns negative in the case where the consumption of the marginal consumer exceeds the average consumption.

The above rules assume that the only basis for distinguishing between consumers lies in the observation of the amount they consume. So in the case discussed above, with a small open-access operator wishing to enter against a large incumbent, one might reasonably conclude that $Q_0$ is small relative to $Q'$; and that $e$ is relatively large. That would mean covering $F$ primarily through markup of $P$ above $c$ rather than markup of $E$ above $v$.

The discussion becomes more interesting when we are able to distinguish between consumers even when they are observed to consume the same amount. Essentially, each consumer type will have a distribution parameter $\theta_i$ and for each type $i$, the critical value will be that of the marginal consumer of that type, $\theta_{i0}$. We would then be able to apply the participation and consumption Ramsey rules above to each of the different consumer types. So as well as balancing participation markup against consumption markup for the entrant, we could also allow for a (small) entrant having a much bigger participation elasticity than a (large) incumbent.

Brown and Sibley cover this by considering ‘increments’ in consumption. A marginal consumer is just willing to pay for an increment in consumption at current consumption levels. All consumers with a stronger taste for the good are more than willing to buy an increment. They then argue that ‘we can think of consumption increments as being bought in “markets” which differ depending on the point from which the increment is made’. Quantity in this case means the number of individuals willing to buy a given increment. They go on to derive a standard Ramsey inverse elasticity rule for pricing increments: the markup of price over marginal cost in each increment market is inversely proportional to the price elasticity of demand in that increment market. Thus, they argue that the nonuniform pricing literature and the uniform pricing literature are united since, correctly interpreted, Ramsey pricing applies in both cases.

So what does this mean for, say, pricing access to a railway network?

We need to locate levels of consumption which are price inelastic and make these contribute as much as possible to the provision of infrastructure and the cost of services. Conversely, price elastic levels of consumption should be priced close to marginal cost. Ramsey nonuniform
pricing looks inside individual markets and sets high markups on ranges of consumption in each market which are relatively inelastic. Ideally, this would mean a higher fixed charge associated with trains with relatively inelastic passenger demand; but if that higher fixed charge led to reductions in peak services, or if indeed it in any way affected marginal decisions, then the benefits of two part tariffs may be lost. If one operated at the level of train operating company, one might there distinguish sufficient differences in the elasticities of participation and consumption between operators to justify different entry fees and different unit charges according to operator. On the other hand, if the operators were competing in the same market, this might well fall foul of the legal requirement of non discrimination.

But overall, the information requirements for effecting an optimal non-linear tariff structure are even greater than achieving an optimal set of Ramsey prices, since we need not only price elasticities but also elasticities of market participation with respect to per-unit charges and access fees together with a distribution of consumer types. On the other hand, if such information is available, then the possibility of multipart tariffs will in general permit a given budget constraint to be met at less welfare cost than under a simple tariff.

There is a further advantage of multipart tariffs where there are specific costs associated with connecting the customer to the network. This is most obvious in the case of markets such as energy, where premises have to be connected to the distribution network. The costs of such connection may then be efficiently charged for in a fixed charge; if the consumer is not willing to pay these costs, then it is efficient for them not to be connected. In transport infrastructure, the presence of the particular customer may impose additional non marginal costs over and above simple connection costs such as connections to private sidings or depots, for instance in a requirement for higher quality track, additional lines etc. Again, such costs may be efficiently recovered via a fixed charge, although the situation becomes complicated if a competitor wishes also to use them. In this case, the entrant would ideally be charged for any loss of surplus towards the fixed charge that their entry imposes on the incumbent.

Pricing of essential facilities

It seems to us that the key question is whether anything can or should be done with the pricing of an essential facility to try and cover the cost of rest of network and/or offset externalities (or other divergences from marginal cost) elsewhere on a network.

The second of these is simpler to address. For example, the road network is largely unpriced at point of use in UK and thus is priced below marginal cost. We can use channel tunnel prices (above MC since it is a complement) to minimise the damage to social welfare along Lipsey/Lancaster lines.

Consider a case as in Britain where the road network is congested but there is no road-pricing scheme or other congestion tax, so that the traveller’s ‘money and time’ costs are less than MSC. What pricing policy should be followed for the (complementary) channel tunnel?

The general rule is

Fel!

(7)
The general implications for pricing $i$ when $j$ is not priced at marginal cost are:

(a) if there is no substitution, set $P_i = MC_i$;

(b) if the goods are substitutes and $P_j > MC_j$, set $P_i > MC_i$.

(c) if the goods are complements and $P_j > MC_j$, set $P_i < MC$.

Note: in a world of more than two interacting goods, the information requirements are likely to be very high and costly.

Thus, in our congested road case, the price of the complementary channel tunnel usage should equal the marginal cost for the channel tunnel plus a markup based on:

(a) the divergence between price and marginal cost of general road use;
(b) the responsiveness of road use to a change in channel tunnel price;
(c) the level of road use relative to channel tunnel use;
(d) the own-price demand elasticity of channel tunnel.

In practice, one might choose to include within the relevant set only those journeys using or potentially using the channel tunnel, that is international trips, for the purpose of generating the relevant elasticities.

The first question seems to us to be just a restricted form of Ramsey pricing, applying the logic (price up on the inelastic link, that is, the essential facility) but essentially ignoring what goes on elsewhere. To the extent that the rest of the network has finite price elasticity of demand which is not included in the relevant calculations, then seeking to recover whole network costs by charging on just part of the network is necessarily inefficient, in that reducing prices on the essential facility and increasing prices on the rest of the network in line with Ramsey rules will more efficiently meet a given budget constraint.

**Ramsey pricing and Peak Loads**

We assume familiarity with the standard Steiner/Boiteux/Williamson treatment and so do not replicate that here. Readers are referred to, for example, Crew and Kleindorfer (1979). We define the unit capacity cost as $\beta$ (the joint cost element) and unit operating cost as $b$ (the specific cost element). We discuss in the context of peak/off-peak although, of course, the treatment applies equally to, say, commuter and long-distance passenger services. Note that what follows assumes constant long run marginal costs, although the analysis can readily be extended.

The purpose of this material is to demonstrate how cost recovery interacts with jointness in supply. In particular, while differential peak/off-peak pricing is efficient when peak/off-peak marginal costs differ, that in itself is not discriminatory pricing. The extensions to second-best financial constraint and profit maximisation clearly show how third-degree discrimination with variable demand requires a proper understanding of true joint/common costs and service-specific costs.
In the firm-peak case, optimal capacity is determined by the peak demand. Assuming perfect divisibility, peak price $P_p$ is set equal to LMC and thus capacity determined as $Q_p$. Once capacity has been thus determined, we have $P_p = \text{LMC} = \text{SMC}$. Given that the capacity has been provided to meet the peak - and paid for solely by the peak users - off-peak demand can be efficiently met by pricing at operating cost. Thus the efficient solution is produced by marginal cost pricing:

$$P_p = b + \beta; \quad P_{op} = b; \quad \text{costs are covered; social surplus is maximised.}$$

Other objectives are possible. Crew and Kleindorfer (1979) provide results for profit maximisation and for meeting a second-best profit constraint. This latter might simply be a target profit level between what would happen with efficient pricing and with profit maximisation. Table 1 presents peak and off-peak prices under different objectives for the firm.

**Table 1: Peak and off-peak prices under different objectives (firm-peak case)**

<table>
<thead>
<tr>
<th>Objective</th>
<th>Peak price</th>
<th>Off-peak price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximise welfare</td>
<td>$b + \beta$</td>
<td>$b$</td>
</tr>
<tr>
<td>Profit–maximising</td>
<td>$\frac{b + \beta}{1 + \frac{1}{\varepsilon_p}}$</td>
<td>$\frac{b}{1 + \frac{1}{\varepsilon_{op}}}$</td>
</tr>
<tr>
<td>Monopolist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second best</td>
<td>$\frac{(1 + \lambda) \cdot (b + \beta)}{1 + \lambda(1 + \frac{1}{\varepsilon_p})}$</td>
<td>$\frac{(1 + \lambda) \cdot b}{1 + \lambda(1 + \frac{1}{\varepsilon_{op}})}$</td>
</tr>
</tbody>
</table>

$\varepsilon_p$ and $\varepsilon_{op}$ are peak and off-peak demand elasticities; $\lambda$ represents the tightness of the budget constraint: $\lambda = 0$ under welfare maximisation and $\lambda \to \infty$ under profit maximisation.

The derivation of the profit-maximising prices is clear. In independent markets, the operator produces where $\text{MR}=\text{MC}$; $b+\beta$ and $b$ are the respective marginal costs; $\text{MR}$ in each case is $P(1 + 1/\varepsilon)$; so $P = \text{MC}/(1 + 1/\varepsilon)$ as stated. The same principle drives the second best case and, when we get to it, the shifting peak variant of this problem; but it is more clearly seen here.

Given that a profit-maximising monopolist will always produce in the region of elastic demand ($\varepsilon < -1$) it can be seen that the monopolist will price above the welfare-maximising levels. This also implies lower output that in the welfare-maximising case. Note that for as long as the off-peak elasticity remains absolutely greater than the peak elasticity at the prevailing prices/demands, off-peak price must always remain below peak price (for positive $\beta$).

It is also clear that pricing to meet a profit constraint higher than the normal economic profit implicit in welfare-maximising will result in higher than efficient prices and lower than efficient output.
In the shifting-peak case, the key difference is that the capacity constraint is binding in both the peak and the off-peak and so both contribute to capacity costs. The share of capital costs borne by each period depends on the tightness of the capacity constraint. In the welfare-maximising case, since the peak by definition has higher willingness to pay than the off-peak, the peak bears a greater share of the capacity costs. In the profit maximising and second best cases it is theoretically possible that the off-peak price exceeds the peak price and so the off-peak may bear a greater share of the capacity cost. In other words, the shapes of the two demand curves are such that there is ‘crossover’ and for some prices, ‘off-peak’ demand exceeds ‘peak’ demand. In such circumstances, it is probably preferable to consider the issue as one of spatial rather than temporal variation in demand.

**Table 2: Peak and off-peak prices under different objectives (shifting-peak case)**

<table>
<thead>
<tr>
<th>Objective</th>
<th>Peak price</th>
<th>Off-peak price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximise welfare</td>
<td>$b + \theta_p$</td>
<td>$b + \theta_{op}$</td>
</tr>
<tr>
<td>Profit–maximising monopolist</td>
<td>$\frac{b + \theta_p}{1 + \frac{1}{\mathcal{E}_p}}$</td>
<td>$\frac{b + \theta_{op}}{1 + \frac{1}{\mathcal{E}_{op}}}$</td>
</tr>
<tr>
<td>Second best</td>
<td>$\frac{(1 + \lambda) \cdot b + \theta_p}{1 + \lambda (1 + \frac{1}{\mathcal{E}_p})}$</td>
<td>$\frac{(1 + \lambda) \cdot b + \theta_{op}}{1 + \lambda (1 + \frac{1}{\mathcal{E}_{op}})}$</td>
</tr>
</tbody>
</table>

$\epsilon_p$ and $\epsilon_{op}$ are peak and off-peak demand elasticities; $\lambda$ represents the tightness of the budget constraint: $\lambda = 0$ under welfare maximisation and $\lambda \rightarrow \infty$ under profit maximisation. $\theta_p$ and $\theta_{op}$ represent the tightness of the capacity constraint in the peak and off-peak respectively.

Let us focus on the second best cases. In the firm peak case, the prices easily rearrange to give a Ramsey pricing rule of the form markup price above marginal cost in inverse proportion to the elasticity, and it is just the notion of marginal cost which changes: operating cost in the off-peak, operating plus capacity cost in the peak.

Interestingly, a similar result obtains for the shifting peak case, where the interpretation of marginal cost is operating cost plus the shadow price of capacity:

$$\frac{P_i - (b + \frac{\theta_i}{1 + \lambda})}{P_i} = \frac{-\lambda \cdot \frac{1}{\mathcal{E}_i}}{1 + \lambda \cdot \mathcal{E}_i}$$

(8)

Thus in a peak load problem, if the elasticities are the same, there is still scope for pricing to be differentiated according to the relative willingness to pay for capacity ($\theta_i$). Conversely, if the willingness to pay for capacity is the same, pricing should still be differentiated according
to elasticity. It is not sufficient in this shifting peak case simply to rely on the price elasticities; joint marginal capacity costs need to be allocated to user groups in accordance with willingness to pay. Note that this is NOT the same as FDC pricing; it is, instead, akin to value-of-service pricing for capacity.

An alternative, developed in the context of airports, but with obvious wider applicability, was proposed by Littlechild and Thompson (1977). It seems to us that the basic Littlechild answer to fair and efficient pricing in the presence of jointness may equally be achieved by a Boiteux/Steiner/Williamson peak load pricing approach with proper specification of marginal costs - and the allocation of joint costs on a WTP basis - along with a Ramsey budget constraint.

Unless the Littlechild approach has been later developed to deal with non-constant costs and/or objectives other than break-even (we are not aware that it has) OR the approach is more readily tractable/amenable to analysis, then we see no particular merit in pursuing it further. It is essentially a cost allocation method with only dubious links to marginality.

Conclusion

The standard Ramsey pricing solution for meeting budget constraints at least cost in terms of welfare is generally applicable to transport infrastructure with some provisos. Firstly, the price charged per vehicle or train kilometre will not simply influence prices charged in the final market but also an important aspect of service quality, namely frequency. This gives added impetus to the use of multipart tariffs, which in any event are generally welfare superior to single part tariffs in the presence of budget constraints. A further advantage of multipart tariffs is that fixed charges may be used to recover any nonmarginal costs specifically attributable to the presence of that customer on the network. Secondly there are constraints on the degree to which price discrimination according to elasticities may be practised by infrastructure managers. One constraint is legal, in that non discrimination rules prevent the charging of different tariffs to competitors in the same market (indeed European rail legislation appears to rule out multipart tariffs in general in such circumstances). Another is practical, in that the infrastructure manager cannot identify and charge separate prices according to final demand characteristics such as ultimate origin and destination, personal characteristics or commodity carried. In general the infrastructure manager will be confined to differentiating on the basis of factors such as section of route, time of day and type of vehicle/train.
References


Lewis, W.A. (1949), *Overhead costs: some essays in economic analysis*, Unwin


Walters, A.A. (1968), *The Economics of Road User Charges*, World Bank


Appendix D. Sharing profit in parallel and serial transport networks
Sofia Grahn-Voorneveld
Sharing profit in parallel and serial transport networks

Sofia Grahn-Voorneveld*

Swedish National Road and Transport Institute
P.O.Box 55685 Stockholm, 102 15 Sweden

Abstract

The pricing schemes for using transport networks differ significantly across countries and modes in Europe. Since the transport networks in different countries usually are connected to each other, the pricing in one country has effects on demand and presumably optimal pricing in other countries. The natural questions arising are whether countries have reason to cooperate with each other when it comes to pricing, and how to handle the profit from such cooperation.

The present paper considers two types of simple networks with congestion; one with parallel links, and one serial network with a number of consecutive links. The owner of each link tolls the traffic using the link. The infrastructure manager of each country tries to maximize a welfare function with respect to the toll.

For these models two types of analysis are performed; First the incentives for cooperative behavior among the countries are studied, and shown to be considerable. This is done by using non-cooperative game theory. Second, cooperative game theory is used to analyse solution concepts for allocating the resources raised from cooperation. For the parallel case the Shapley value is proven to have very nice properties, such as being the bary-centre of the core. For the serial case the Shapley value is shown not to be a good option. Instead three new allocation rules are introduced.

Keywords: Transport networks, game theory, cooperative game theory.
JEL codes: C71; C72; H71; L92; R41.

*I thank Gunnar Lindberg, Jan-Eric Nilsson, Mark Voorneveld and seminar participants for useful comments.
1 Introduction

In a transport network it is often the case that different links are owned by, or under the jurisdiction of different countries, regions or companies. A lot of attention has been given to the principle of marginal cost pricing. However, the revenue from such pricing may not be sufficient to make up for the full costs of the transport network. Nor will it provide resources for new infrastructure investments. It is therefore of interest to consider other pricing principles.

The pricing schemes for using the transport networks differ significantly across countries and modes in Europe. Since transport networks in different countries often are connected to each other, the pricing in one country has effects on demand and presumably optimal pricing in other countries. Natural questions arising are whether countries have reason to cooperate with each other when it comes to pricing, and how to handle the profit from such cooperation.

The issue of optimal pricing in transport networks is not new. There are a number of studies on parallel network structures studying various aspects of pricing of parallel congestible roads, see for instance Braid (1986), Verhoef et al. (1996), De Palma and Lindsey (2000), McDonald and Liu (1999), Small and Yan (2001), van Dender (2005), and de Borger et al. (2006). There are also several studies addressing pricing in serial networks, see for instance Levinson (2001), de Borger et al. (2006), Bassanini and Pouyet (2005) and Agrell and Pouyet (2006). However, none of these studies consider cooperative behavior among the owners of a network.

The purpose of this paper is first, to show what incentives there are for cooperative behavior among countries together owning a transport network, and second, to provide insights and tools for cooperative behavior among the owners of different parts of both parallel and serial networks. The model is similar to the model of de Borger et al. (2006) although in the present paper capacity investments are not considered.

The present paper considers two types of simple networks with congestion:

- A parallel network with a number of parallel links. The links are considered to be substitutes, meaning that transit traffic chooses between the alternative parallel routes: local traffic can only use the local link.

- A serial network with a number of consecutive links, together forming a transport corridor. Transit traffic by definition passes through every link, whereas local traffic only uses the local link.

The owner of each link tolls the traffic using the link, and the infrastructure manager of each country tries to maximize a welfare function with respect to the toll. The infrastructure manager only considers the welfare of his/her own country.

The parallel model capture situations like competing main routes through a continent or the transalpine crossings. The serial model captures situations like serial sections of the Trans European Networks and interstate highways in the US.
In the parallel case the links are substitutes. It is therefore expected that cooperation leads to higher tolls on transit traffic and higher welfare for the owners of the network. In the serial case it turns out that cooperation is not only beneficial for the owners of the network but also for the users, since cooperation will in fact reduce the tolls. Although this result is not new, the structure of optimal toll or tax structure has not been fully investigated. The previous models analyze serial networks with two consecutive links. By modelling networks with \( n \) links in a network this paper can provide a more thorough analysis of toll structure and welfare effects. It is shown that infrastructure investments and maintenance in one country affects the welfare also in other countries, implying that not only pricing, but also decisions concerning infrastructure investments and maintenance made on local level might be inefficient with respect to the total welfare level.

For cooperation to occur it is not enough that the total welfare level increases compared to non-cooperation, the countries also have to be able to agree on how to split the resources raised from cooperation. The analysis shows that this cannot be done via a uniform toll level and each country keeping their own toll. In the parallel case this is obvious since in equilibrium the total user cost including tolls must be equal for every link, and since other user costs vary, so must the tolls. In the serial case it is unreasonable to expect a country with a very costly link to accept setting the same toll as a country with a less costly link.

Instead of setting a uniform toll, the total income from cooperation has to be allocated among the cooperating countries. Of course no "correct" such allocation exist; however, some allocations are more likely to be accepted than others. These are allocations that satisfy intuitive properties related to fairness. By supplying such rules, negotiation costs are reduced and cooperation is more likely to occur.

The second type of analysis performed in the paper deals with such allocations for the serial and parallel transport models. To be able to analyze the cooperative situation thoroughly, cooperative game theory is used. For this purpose a new class of problems is introduced - transport network problems. Both the serial and the parallel model fit in this class of problems. Further a new class of cooperative games is introduced - the class of parallel transport network games. This class of games corresponds to problems like the parallel model.

The Shapley value is one of the most well-known solution concepts in cooperative game theory. In the parallel case it is easy to motivate the use of the Shapley value since it has very nice properties for this class of games, it is proven to coincide with the bary-center of the core. In the serial case however, the Shapley value coincides with setting a uniform toll level and letting each country keep their toll incomes. As mentioned above this is not a reasonable allocation. Instead, three new allocation rules are introduced. These rules are inspired by rules from the literature on bankruptcy games. See for instance Thomson (2003).

The set-up of this paper is as follows. In section 2 the basic model is introduced. Section 3 analyzes the toll level and welfare level in the serial case with and without cooperation. In section 4 cooperative game theory is used to analyze allocation rules for the serial model. In section 5 the parallel case is analyzed with respect to optimal toll
and welfare levels, with and without cooperation. Section 6 studies allocation rules for the parallel model. Section 7 deals with the case when there is a demand both for transit traffic and local traffic. Finally, section 8 concludes.

2 The model

Two types of networks are considered. The parallel network arises when there are a number of competing transport corridors through different countries, meaning that the countries compete for toll/tax revenue. The serial network arises when a monopolistic transport corridor runs through a number of sequential countries or regions.

Each country owns one arc of the network and is allowed to toll traffic through this link. The traffic consists of local traffic passing through one link only and with no other options, and transit traffic which in the parallel case can choose between all parallel links, and in the serial network passes through all sequential links. As opposed to the model of de Borger et al. (2006), which deals only with two countries, this model is generalized to $n$ countries. This allows a more thorough analysis to be performed.

Let $N = \{1, \ldots, n\}$ be the set of countries/owners of the network. Each member of $N$ owns one link of the network.
Let \( t_i \) be the toll on local transport, and \( \tau_i \) the toll on transit traffic in country \( i \), all \( i \in N \). Demand for local and transit traffic is represented by the strictly decreasing and twice differentiable inverse demand functions

\[
\begin{align*}
p_i^l(x_i(t_i)) & \quad \text{local traffic } \forall i \in N \\
p^t(x(\tau)) & \quad \text{transit traffic}
\end{align*}
\]

where \( x(\tau) \) is the demand for transit traffic and \( x_i(t_i) \) is the demand for local traffic in country \( i \) for all \( i \in N \). The inverse demand functions are generalized prices including time costs and tolls for all \( i \in N \). Let \( r_i \) be the user resource- plus time cost on arc \( i \). Due to congestion this cost depends on how much traffic there is on link \( i \), so \( r_i \) is a function of transit traffic \( x(\tau) \) plus local traffic \( x_i(t_i) \).

We have to separate the parallel and the serial case since transit traffic goes through all arcs in the serial case. For the parallel case let \( y_i(\tau) \) be the demand for transit traffic through path \( i \) for all \( i \in N \), with \( \sum_{i=1}^{n} y_i(\tau) = x(\tau) \).

The generalized user cost for local traffic travelling through arc \( i \) is for all \( i \in N \) given by

\[
g_i^l = r_i(x(\tau) + x_i(t_i)) + t_i \quad \text{for the serial case},
\]

\[
g_i^p = r_i(y_i(\tau) + x_i(t_i)) + t_i \quad \text{for the parallel case},
\]

and the generalized user cost for transit traffic travelling through arc \( i \) is for all \( i \in N \) given by

\[
g_i^l = r_i(x(\tau) + x_i(t_i)) + \tau_i \quad \text{in the serial case},
\]

\[
g_i^p = r_i(y_i(\tau) + x_i(t_i)) + \tau_i \quad \text{in the parallel case}.
\]

In equilibrium the generalized user cost equals the generalized user price for both local and transit traffic. Thus for all \( i \in N \) we have that

\[
\sum_{i=1}^{n} g_i^l = \sum_{i=1}^{n} \left[ r_i(x(\tau) + x_i(t_i)) + \tau_i \right] = p^t(x(\tau)) \quad \text{for transit traffic in the serial case}
\]

The infrastructure manager in each country maximizes social welfare consisting of consumer surplus for local users, tax revenue and maintenance cost. The infrastructure manager considers only the welfare of his/her own country. For each country \( i \in N \) the welfare function is given by:

\[
W_i = \int_0^{x_i} p_i(\tilde{x}_i)d\tilde{x}_i - g_i^l(x(\tau) + x_i(t_i)) + t_i x_i(t_i) + \tau_i x(\tau) - c_i \quad \text{for the serial case} \quad (1)
\]

\[
W_i = \int_0^{x_i} p_i(\tilde{x}_i)d\tilde{x}_i - g_i^p(y_i(\tau) + x_i(t_i)) + t_i x_i(t_i) + \tau_i y_i(\tau) - c_i \quad \text{for the parallel case} \quad (2)
\]
where $c_i$ is the maintenance cost for the road segment in country $i$.

The welfare functions show that it is not possible to say, for the general case, whether tolls equal to marginal cost pricing cover the costs or not.

### 3 Welfare levels with and without cooperation in the serial case

Assume that all cost and demand functions are linear. Further assume a simple case where local demand is zero, thus reducing the welfare function to $W_i(\tau) = \tau_i x(\tau) - c_i$ for all $i \in N$.

$$p^t(x) = a - bx(\tau) \text{ with } a, b > 0$$
$$g^t_i = r_i(x(\tau)) + \tau_i = \alpha_i + \beta_i x(\tau) + \tau_i \text{ with } \alpha_i, \beta_i > 0 \text{ for all } i \in N$$

this gives that

$$\sum_{i=1}^{n} g^t_i = \sum_{i=1}^{n} [\alpha_i + \beta_i x(\tau) + \tau_i] = a - bx(\tau)$$

which gives the reduced demand function

$$x(\tau) = \frac{a - \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \tau_i}{b + \sum_{i=1}^{n} \beta_i} \tag{3}$$

We assume that $a > \sum_{i=1}^{n} \alpha_i$ i.e. that there is a demand for transit traffic. Let

$$A = a - \sum_{i=1}^{n} \alpha_i$$
$$B = \frac{1}{b + \sum_{i=1}^{n} \beta_i}$$

and rewrite (3) as

$$x(\tau) = B(A - \sum_{i=1}^{n} \tau_i) \tag{4}$$

The welfare function (1) now reduces to

$$W_i(\tau) = \tau_i B(A - \sum_{i=1}^{n} \tau_i) - c_i \text{ for all } i \in n. \tag{5}$$
In order to calculate the toll-level of each country in equilibrium non-cooperative game theory is used. A game is a triplet $\langle N, T, W \rangle$ where $N = \{1, ..., n\}$ is the set of players (in this case the set of infrastructure managers), and for each player $i$ a set of strategies is given by $T_i$. The set of strategies of a player $i$ here consists of the set of possible toll levels. The set of all possible strategy profiles of the players is given by $T = \times_{i \in N} T_i$. For each player $i$ and strategy profile $\tau \in T$, the function $W_i(\tau)$ specifies the welfare level of player $i$, in this model given by function (5).

In order to find the Nash equilibrium of the game corresponding to the serial model, we need to calculate the best response function for every player $i$, i.e. the the toll level that maximizes the welfare of country $i$ given the toll levels in all other countries. The best response function of infrastructure manager $i$ is given by the first-order condition

$$\frac{dW_i(\tau)}{d\tau_i} = B \left( A - \sum_{j=1}^{n} \tau_j - \tau_i \right) = 0$$

which gives the optimal toll

$$\tau_i^* = A - \sum_{i=1}^{n} \tau_i \text{ for all } i \in N$$

Since this is the case for all $i \in N$ the optimal toll level $\tau_i^*$ is equal for all countries, hence

$$\begin{align*}
\tau_i^* &= A - n\tau_i^* \\
&= \frac{A}{n+1} \forall i \in N 
\end{align*}$$

This is the unique Nash equilibrium of the problem. Note that the toll level of each country is independent of the congestion parameter $\beta_i$, hidden in $B$, for all $i \in N$. This is a consequence of the serial structure and the linearity of the inverse demand function and the generalized user cost functions, which cause the resulting demand function for transit traffic (4) to be a linear function of the total toll level $\sum_{i=1}^{n} \tau_i$, rather than a function of the individual toll levels of the countries in the network.

The total toll level on the user will be the sum of the toll in every country:

$$\sum_{i=1}^{n} \tau_i = \frac{nA}{n+1}$$

It is interesting to see that the optimal toll level is a function of the number of countries owning the transport corridor. The total toll level is close to $A$ when the transport corridor is owned by a large number of countries $n$, resulting in a demand for transit traffic close to zero.

Looking at the demand function (3) it is reasonable to limit the toll $\tau_i$ to the interval $\tau_i \in [0, a - \sum_{i=1}^{n} \alpha_i]$. Since $W_i$ is a continuous function and we optimize over a compact set there exist a maximum. Checking the second order condition show that the first order condition gives a maximum.
Inserting the optimal toll (6) into the demand function (3) gives

\[ x^* = B(A - \frac{nA}{n + 1}) = \frac{AB}{n + 1} \]

which gives the welfare

\[ W_i^* = \frac{A^2B}{(n + 1)^2} - c_i \text{ for all } i \in N \]

\[ \sum_{i=1}^{n} W_i^* = \frac{nA^2B}{(n + 1)^2} - \sum_{i=1}^{n} c_i \] (8)

Assume instead that the infrastructure managers of the \( n \) countries cooperate to maximize the total welfare. Using (5) the total welfare can be written as:

\[ \sum_{i=1}^{n} W_i(\tau) = B(A \sum_{i=1}^{n} \tau_i - \sum_{i=1}^{n} \tau_i \sum_{j=1}^{n} \tau_j) - \sum_{i=1}^{n} c_i \] (9)

The first order condition\(^2\)

\[ \frac{\partial}{\partial \tau_i} \sum_{j=1}^{n} W_j(\tau) = B(A - \sum_{j=1}^{n} \tau_j - \sum_{j=1}^{n} \tau_j) = 0 \text{ } \forall i \in N \]

gives

\[ A - 2 \sum_{i=1}^{n} \tau_i = 0 \]

\[ \Rightarrow \sum_{i=1}^{n} \tau_i^{\text{coop}} = \frac{A}{2} \] (10)

Note that this is the optimal total toll level, the individual toll \( \tau_i \) is not specified. It is intuitive that the demand for transit traffic, given the total toll level, is independent of the individual toll.

Inserting (10) in to the demand function (3) gives

\[ x^{\text{coop}} = B(A - \frac{A}{2}) = \frac{AB}{2} \text{ for all } i \in N \]

which inserted in (9) results in the total welfare

\[ \sum_{i=1}^{n} W_i^{\text{coop}} = \frac{AB A}{2} - \sum_{i=1}^{n} c_i = \frac{A^2B}{4} - \sum_{i=1}^{n} c_i \] (11)

\(^2\)Since the toll \( \tau_i \) is limited to the interval \( \tau_i \in [0, a - \sum_{i=1}^{n} \alpha_i] \) for all \( i \in N \), and \( \sum_{i=1}^{n} W_i \) is a continuous function and we optimize over a compact set there exist a maximum. Using the Kuhn-Tucker conditions shows that the first order condition gives a maximum.
Comparing the total welfare level of cooperation (11) and non-cooperation (8) show that

\[ \sum_{i=1}^{n} W_{i}^{coop} \geq \sum_{i=1}^{n} W_{i}^{*} \text{ with strict inequality for } n > 1 \]

since

\[ \frac{A^2 B}{4} - \sum_{i=1}^{n} c_i \geq \frac{n A^2 B}{(n+1)^2} - \sum_{i=1}^{n} c_i \]

\[ \Rightarrow \frac{1}{4} \geq \frac{n}{(n+1)^2} \text{ with strict inequality for } n > 1. \]

Cooperation is thus Pareto dominant. Note that the special case \( n = 1 \) gives equality. This is trivial since cooperation and non-cooperation are the same thing when the transport corridor is owned by one country.

### 3.1 Comparing the implications of cooperation and non-cooperation

A comparison between the total toll levels of cooperation (10) and non-cooperation (7) shows that cooperation reduces total toll level for the users.

\[ \sum_{i=1}^{n} \tau_{i}^{coop} \leq \sum_{i=1}^{n} \tau_{i}^{*} \]

\[ \frac{A}{2} \leq \frac{n A}{n+1} \text{ with strict inequality for } n > 1. \]

For a large number of links in the transport corridor, the total toll level in the non-cooperative case is close to \( A \), i.e. twice the total toll level in the case of cooperation. Note that the demand for transit traffic becomes zero when the total toll level is \( A \). The reason why the tolls are higher in the non-cooperative case than the cooperative case is that a country in the case of non-cooperation receives all profit from rising his own toll, but that the negative effects - decreased demand for transit traffic - affect all countries among the serial transport corridor equally. In the literature this situation is often compared to the problem of vertical integration in a supply chain, where overall markups are higher but total profit lower than in the case with full integration. The situation also shows similarities to what in welfare economics is called the ”tragedy of the commons”. The owners of the serial network own the transport corridor together but since the profit is individual and the cost shared with the rest of the owners, the corridor will be overexploited with respect to tolls.

A similar problem can be seen for the user cost parameters \( \alpha_i \) and \( \beta_i \) (hidden in \( A \) and \( B \) above). Remember that

\[ A = a - \sum_{i=1}^{n} \alpha_i \text{ giving that } \frac{dA}{d\alpha_i} = -1 \]
\[ B = \frac{1}{(b + \sum_{i=1}^{n} \beta_i)} \] giving that \[ \frac{dB}{d\beta_i} = -B^2 \]

Derivating the welfare level for the non-cooperative case, \( W_j^* \), with respect to \( \alpha_i \) and \( \beta_i \) respectively gives

\[
\frac{dW_j^*}{d\alpha_i} = \frac{2AB}{(n + 1)^2} \frac{dA}{d\alpha_i} = -\frac{2AB}{(n + 1)^2} < 0 \text{ for all } i, j \in N
\]

\[
\frac{dW_j^*}{d\beta_i} = \frac{A^2}{(n + 1)^2} \frac{dB}{d\beta_i} = -\frac{A^2B^2}{(n + 1)^2} < 0 \text{ for all } i, j \in N
\]

which shows that a change in \( \alpha_i \) and \( \beta_i \) affects all countries equally with a reduction in toll income and welfare. The total welfare reduction will be \( n \) times larger than the individual welfare reduction. An infrastructure investment in country \( i \) that will reduce \( \alpha_i \) and/or \( \beta_i \) can be inefficient for country \( i \) but efficient with respect to the total welfare level. The infrastructure investments will therefore be smaller in the case of non-cooperation than with cooperation.

Although this model views \( \alpha, \beta \) and \( c \) as parameters, it is logical that a reduction of maintenance costs \( c_i \) increases the user cost of the road segment \( i \), i.e. increasing the user cost parameters \( \alpha_i \) and/or \( \beta_i \) or that investments in segment \( i \) reduces \( \alpha_i \) and/or \( \beta_i \).

This implies that not only pricing but also decisions concerning maintenance and infrastructure investments, made on local level, might be inefficient concerning the total welfare level. Without regulation or cooperation the tolls would be higher than what is efficient from the users and the infrastructure suppliers’ viewpoint. Further the standard of infrastructure would be lower than what is efficient.

Countries not owning links in the serial network are not included in the welfare analysis. However, it is obvious that also their welfare will improve by cooperation among the countries who do own links in the network.

### 4 Sharing profit in the serial case

For cooperation to occur it is not enough that the total welfare level increases compared to non-cooperation, the countries also have to be able to agree on how to split the resources raised from cooperation. One way to do this is to set a uniform toll level \( \tau_i = \frac{1}{n} \sum_{i=1}^{n} \tau_i^{coop} = \frac{A}{2n} \) and let every country keep keep their toll income. In the serial model this results in the individual welfare

\[ W_i = \frac{A^2B}{4n} - c_i \text{ for all } i \in N. \]
In practice this means that a country with a long road with many tunnels and bridges could charge no more than a country with a short and relatively cheap road. First of all this might seem unfair, second the fact that countries with high costs do not get higher profit can cause problems concerning maintenance level and investments in infrastructure. If a country with a high maintenance cost cuts down on maintenance in order to increase its welfare, this will increase $\alpha_i$ and/or $\beta_i$. Even if this increases the welfare of country $i$, it might very well reduce the total welfare. The division of the profit from cooperation is therefore of great importance.

To be able to analyze the cooperative situation thoroughly, cooperative game theory is used. A cooperative game is a tuple $(N, v)$ where $N = \{1, \ldots, n\}$ is the set of players (in this case the set of infrastructure managers of countries owning links in the serial network). The set of all possible coalitions of players in $N$ is denoted by $2^N = \{S \mid S \subseteq N\}$. The function $v : 2^N \to \mathbb{R}$ is called the characteristic function of the game, and assigns to each coalition $S$ a value $v(S) \in \mathbb{R}$, with $v(\emptyset) = 0$.

A common interpretation of a cooperative game, or a coalitional form game, is that it models a situation in which the actions of players who are not part of coalition $S$ do not influence the value $v(S)$, i.e. that the players in coalition $S$ can act isolated from the rest of the players in $N$. In the model discussed here this is not the case. All $n$ countries are needed to form the transport corridor. Further it is obvious that welfare of a coalition $S$ depends on the toll level chosen by the players outside $S$, since their toll influence the demand for transit traffic.

For situations like this the characteristic function is usually derived from a problem in one of two ways. The first is based on what a player or coalition can guarantee himself/themselves when the remaining players act to minimize their payoff (in this context the highest welfare level $\sum_{i \in S} W_i$ coalition $S$ can guarantee themselves if they have to reveal $\sum_{i \in S} \tau_i$ and the remaining players, knowing this, choose $\sum_{i \in N \setminus S} \tau_i$ in order to minimize the welfare of coalition $S$), the second is based on the payoff to which the remaining players can hold a coalition (in this context the highest welfare level $\sum_{i \in S} W_i$ coalition $S$ can guarantee themselves when the players outside $S$ choose $\sum_{i \in N \setminus S} \tau_i$ in order to minimize $\sum_{i \in S} W_i$ before coalition $S$ make their choice of $\sum_{i \in S} \tau_i$). For more details see for instance Friedman (1991).

Looking at the demand function (4) $x(\tau) = B(A - \sum_{i=1}^{n} \tau_i)$ it is easy to see that a toll level $\sum_{i \in N \setminus S} \tau_i = A$ for the players outside $S$ will yield coalition $S$ the welfare level $\sum_{i \in S} W_i = \sum_{i \in S} c_i \leq 0$ for both approaches. This behavior of $N \setminus S$ is however not credible since also $\sum_{i \in N \setminus S} W_i = -\sum_{i \in N \setminus S} c_i \leq 0$. This is a common problem for situations where all players are needed to achieve something. The way to handle this is to define a reasonable characteristic function specially for this case.

For our serial case we have that when the countries do not cooperate they get the
welfare

\[ W_i^* = \frac{A^2B}{(n+1)^2} - c_i \text{ for all } i \in N \]

which consists of a constant \( \frac{A^2B}{(n+1)^2} \), equal for every country, minus the individual cost parameter \( c_i \). Therefore it seems reasonable to define the value function \( v \) in the same fashion. Let

\[ v(S) = z_{|S|} - \sum_{i \in S} c_i \text{ for all } S \in 2^N \setminus \emptyset \]

and \( v(\emptyset) = 0 \)

where

\[ z_{|N|} = z_n = \frac{A^2B}{4} = \sum_{i=1}^{n} W_i^{coop} + \sum_{i=1}^{n} c_i \]

and \( z_{|S|} \) is a constant in \( \mathbb{R}_+ \), equal for all coalitions with the same number of members.

There exist a number of solution concepts in cooperative game theory, both concepts which result in a set of allocations and concepts resulting in a unique allocation for every game. In practice a concept pointing out a unique solution is of course appealing. The most well established such solution concept is the Shapley value. The payoff to each player is the average marginal contribution that the player makes to each of the coalitions to which he belongs.

The Shapley value was introduced by Shapley (1953) and can be characterized by four axioms. Somewhat informally:

- **efficiency**, i.e. no resources are wasted;
- **anonymity**, i.e. two identical players are treated equally;
- **dummy property**, i.e. a player with a constant marginal contribution to every coalition of which he is a member, is allocated this constant.
- **additivity**, i.e. the solution of the sum of two games, is the sum of the solution to the two games.

The Shapley value is the only solution concept satisfying all four of these axioms.

**Example 4.1** Let \( \langle N, v \rangle \) be a 3-person game with

\[
\begin{align*}
v(\{i\}) &= 0 \text{ for all } i \in \{1, 2, 3\} \\
v(\{1, 2\}) &= 4 \\
v(\{1, 3\}) &= 7 \\
v(\{2, 3\}) &= 15 \\
v(N) &= 20
\end{align*}
\]
For each permutation \( \sigma \) of the members of \( N \) we get a vector of marginal contributions. Let \( m_i^\sigma(v) \) denote the marginal vector of player \( i \). The payoff vector corresponds to a situation where the players enter a room one by one in the order \( \sigma(1), \sigma(2), \ldots, \sigma(n) \) and where each player is given the marginal contribution he creates by entering.

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( m_1^\sigma(v) )</th>
<th>( m_2^\sigma(v) )</th>
<th>( m_3^\sigma(v) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2,3)</td>
<td>0</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>(1,3,2)</td>
<td>0</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>(2,1,3)</td>
<td>4</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>(2,3,1)</td>
<td>5</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>(3,1,2)</td>
<td>7</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>(3,2,1)</td>
<td>5</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td><strong>sum</strong></td>
<td><strong>21</strong></td>
<td><strong>45</strong></td>
<td><strong>54</strong></td>
</tr>
</tbody>
</table>

The average of the six marginal vectors is \( \frac{1}{6} (21, 45, 54) \), which by definition is the Shapley value of the game \( (N, v) \).

Let \( \Phi_i(v) \) denote the Shapley value of a game \( (N, v) \), then the Shapley value can be written as

\[
\Phi_i(v) = \sum_{S:i \notin S} \frac{|S|!(n - 1 - |S|)!}{n!} (v(S \cup \{i\}) - v(S)) \tag{12}
\]

Using the characteristic function defined above gives

\[
\Phi_i(v) = \sum_{S:i \notin S} \frac{|S|!(n - 1 - |S|)!}{n!} (v(S \cup \{i\}) - v(S)) = \sum_{S:i \notin S} \frac{|S|!(n - 1 - |S|)!}{n!} \left( z_{|S\cup\{i|} - \sum_{j \in S \cup \{i\}} c_j - z_{|S|} + \sum_{j \in S} c_j \right) = \sum_{S:i \notin S} \frac{|S|!(n - 1 - |S|)!}{n!} (z_{|S\cup\{i|} - z_{|S|} - c_i) \tag{13}
\]

The number of coalitions, with \( k \) members, that can be formed from the set \( N \setminus \{i\} \) is \( \frac{(n-1)!}{k!(n-1-k)} \) using this we can rewrite (13) as

\[
\Phi_i(v) = \sum_{k=1}^{n-1} \frac{(n-1)!}{k!(n-1-k)} \frac{k!(n-1-k)!}{n!} (z_{k+1} - z_k - c_i) + \frac{1}{n} (z_1 - c_i) = \frac{1}{n} \sum_{k=0}^{n-1} (z_{k+1} - z_k - c_i) = \frac{1}{n} (z_n - nc_i)
\]
Thus in the serial case the Shapley value coincides with splitting the profit from cooperation equally. As discussed above this division has major shortcomings. Therefore the Shapley value is not as appealing for this type of problem as it is for many other situations. Instead, other options can be constructed specifically for this type of problems. For this purpose a new class of problems is defined below.

**Definition 4.2** A *transport network problem* is a tuple \((N, P, c, W)\) consisting of

- a finite set of network owners \(N \subset \mathbb{N}\);
- a profit \(P\) from cooperation (in the serial model \(P = \sum_{i=1}^{n} W_i^{coop} - \sum_{i=1}^{n} W_i^*\));
- a function \(c : N \to \mathbb{R}_+\) specifying the cost of each network owner;
- a function \(W : N \to \mathbb{R}_+\) specifying the welfare from non-cooperation for each member of \(N\).

Let \(T\) denote the set of all transport network problems. A transport network rule is a function \(\varphi\) on \(T\), that assigns to each serial network problem \((N, P, c, W) \in T\) a vector \(\varphi(N, P, c, W) \in \mathbb{R}^N\), specifying payoffs to each member of \(N\). Note that this class of problems apply for both the serial and the parallel case, but is even more general than our models. The network structure is not specified. Further the only specification of the welfare function is that it should be of the form \(W_i = f_i^* - c_i\), where \(f_i^* \in \mathbb{R}_+\) is viewed as an individual constant. For the specific model of section 3 we can rewrite the welfare function as \(W_i = f_i(\tau) - c_i\) for all \(i \in N\), then \(f_i^* = f_i(\tau^*)\).

Below three new allocation rules are formulated for this type of problems.

**Definition 4.3**

- The *proportional rule* \(PR\) assigns to each \((N, P, c, W) \in T\) and \(i \in N\) the amount
  \[
  PR_i (N, P, c, W) = W_i^* + P \frac{c_i}{\sum_{j=1}^{n} c_j}
  \]
  that is the profit from cooperation is divided proportionally to the cost of each country.

---

\[\begin{align*}
= & \frac{1}{n} \sum_{i=1}^{n} c_i \\
= & \frac{A^2B}{4n} - c_i
\end{align*}\]
• The *adjusted proportional rule* APR assigns to each \((N, P, c, W) \in T\) and \(i \in N\) the amount

\[
APR_i (N, P, c, W) = W_i^* + P \left( \frac{1-\lambda}{n} \right) + P\lambda \sum_{j=1}^{n} \frac{c_i}{c_j}
\]

where \(\lambda \in (0, 1)\)

that is part of the profit \(P (1 - \lambda)\) is divided equally among the players and the rest \(P\lambda\) is divided proportionally to the cost of each country.

• The *adjusted equal profit rule* AP assigns to each \((N, P, c, W) \in T\) and \(i \in N\) the amount

\[
AP_i (N, P, c, W) = W_i^* + c_i + \frac{P - \sum_{j=1}^{n} c_j}{n}
\]

that is each country is first compensated for its costs, thereafter the rest of the profit is split equally among the countries.

Consider two countries with \(\alpha_i = \alpha_j, \beta_i = \beta_j\) and \(c_i > c_j\), then all three rules allocate more to player \(i\) then to player \(j\), which was not the case with the Shapley value.

There are some intuitive properties that seem reasonable to demand from an allocation rule for the class of transport network problems:

**Definition 4.4** A transport network rule \(\varphi\) is

• **efficient** if \(\sum_{i=1}^{n} \varphi_i (N, P, c, W) = P + W_N\) for all \((N, P, c, W) \in T\);

• **individually rational** if \(\varphi_i (N, P, c, W) \geq W_i\) for all \(i \in N\) and all \((N, P, c, W) \in T\);

• **P-monotonic** if for each pair \((P, N, c, W), (P', N, c, W) \in T\) with \(P' \geq P\) we have that \(\varphi_i (N, P', c, W) \geq \varphi_i (N, P, c, W)\) for all \(i \in N\);

• **W-monotonic** if for each pair \((P, N, c, W), (P, N, c, W') \in T\) with \(W_i \geq W_i'\) we have that \(\varphi_i (N, P, c, W') \geq \varphi_i (N, P, c, W)\).

An efficient transport network rule divides the total welfare over the cooperating countries. An individually rational rule gives each country at least what they would have gotten in the case of non-cooperation. If the profit from cooperation increase, P-monotonicity implies that each country get at least as much as before. W-monotonicity means that if a country increase its welfare in the case of non-cooperation, this country will get at least as much as before even if the profit from cooperation does not increase.

It is trivial to see that the rules introduced above are efficient, individually rational, P-monotonic and W-monotonic for the whole class of transport network problems.
The Shapley value is defined as an allocation rule for the class of cooperative games, not as a transport network rule for the class of transport network problems. However, it is possible to rewrite the Shapley value as a transport network rule for the serial model. The Shapley value then becomes

\[
\Phi_i = \frac{A^2 B}{4n} - c_i = \frac{P}{n} + W_i^* \tag{14}
\]

Note the similarity to the APR-rule if \( \lambda = 0 \), i.e. when nothing of the profit from cooperation is split proportional to the cost of each country. Using the expression (14) it is easy to see that also the Shapley value is efficient, individually rational, \( P \)-monotonic and \( W \)-monotonic for the serial model. However, this is not true for the general case. Using the definition (12) for the Shapley value and the expression for individual rationality used in the literature on cooperative game theory we have that a solution concept \( f \) is individually rational if \( f_i(v) \geq v(\{i\}) \) for all characteristic functions and all \( i \in N \). Consider the following example:

**Example 4.5** Let \( \langle N, v \rangle \) be a cooperative game with player set \( N = \{1, 2, 3\} \) and characteristic function

\[
\begin{align*}
v(\{i\}) &= 1 \text{ for all } i \in N \\
v(\{1, 2\}) &= v(\{1, 3\}) = 0 \\
v(\{2, 3\}) &= 5 \\
v(N) &= 6
\end{align*}
\]

Then the vectors of marginal contributions of the players for each permutation are given by

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( m_1^2(v) )</th>
<th>( m_2^2(v) )</th>
<th>( m_3^2(v) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2,3)</td>
<td>1</td>
<td>-1</td>
<td>6</td>
</tr>
<tr>
<td>(1,3,2)</td>
<td>1</td>
<td>6</td>
<td>-1</td>
</tr>
<tr>
<td>(2,1,3)</td>
<td>-1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>(2,3,1)</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>(3,1,2)</td>
<td>-1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>(3,2,1)</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td><strong>sum</strong></td>
<td><strong>2</strong></td>
<td><strong>17</strong></td>
<td><strong>17</strong></td>
</tr>
</tbody>
</table>

The Shapley value is the average of the marginal vectors resulting in \( \left( \frac{1}{3}, \frac{17}{6}, \frac{17}{6} \right) \). The Shapley value allocates the value \( \frac{1}{3} \) to player 1 although \( v(\{1\}) = 1 \), showing that the Shapley value is not individually rational for the general case. \( \square \)

\(^3\)Note that this expression for the Shapley value is true only for the serial model, not for the whole class of transport network problems.
Using the expression (14) of the Shapley value it is easy to see that
\[
\frac{\partial \varphi_j(N, P, c, W)}{\partial P} \frac{\partial P}{\partial A} \frac{\partial A}{\partial \alpha_i} < 0 \quad \text{for all } j \in N \text{ and } c_j > 0
\]
\[
\frac{\partial \varphi_j(N, P, c, W)}{\partial P} \frac{\partial P}{\partial B} \frac{\partial B}{\partial \beta_i} < 0 \quad \text{for all } j \in N \text{ and } c_j > 0
\]
for the PR, APR, AR rules and the Shapley value for the serial model. This shows that the individual user cost parameters of a country affects the profit of all countries. The user cost parameter \( \beta_i \) captures all costs related to congestion and \( \alpha_i \) all other costs of road segment \( i \). An infrastructure investment that will lower \( \alpha_i \) and/or \( \beta_i \) will thus increase the welfare in all countries in \( N \) and not only in country \( i \). It is possible that such an investment is not profitable for country \( i \), and in a non-cooperative situation the investment would not be made. However in the cooperative situation it is easy to see that the investment might be beneficial if the total welfare gain of the countries along the transport corridor is considered. By using some of the profit from cooperation for contributions to such national investments, everyone can increase their welfare, including the users.

For the PR, APR and the AP rule it is obvious that a country has incentive to claim larger costs than is the case. One way to handle this problem is to divide the transport corridor into segments considered roughly equally costly, and that \( c_i \) is exchanged to a constant times the number of such segments owned by \( i \). This way the countries do not have to report their costs every time the profit is to be divided.

All three of the new allocation rules introduced above have very nice properties for the serial problem, and is therefore recommended as allocationa rules for problems of this type.

5 Welfare levels with and without cooperation in the parallel case

As in the sequential case we assume linear demand and cost functions. Further assume that the local demand is zero. The total demand \( x \) for transit transport is the sum of the demand for transit transport through all parallel paths in the network.

\[
p^i(x(\tau)) = a - bx(\tau) = a - b \sum_{i=1}^{n} y_i(\tau) \quad \text{with } a, b > 0
\]
\[
g^i_i = r_i(y_i(\tau)) = \alpha_i + \beta_i y_i(\tau) + \tau_i \quad \text{with } \alpha_i, \beta_i > 0 \quad \text{for all } i \in N
\]
where \( y_i \) is the demand for transit traffic through path \( i \). To simplify the calculations we assume that \( \alpha_i = \alpha \), and \( \beta_i = \beta \) for all \( i \in N \). This gives

\[
g^i_i = r_i(y_i(\tau)) = \alpha + \beta y_i(\tau) + \tau_i \quad \text{with } \alpha, \beta > 0 \quad \text{for all } i \in N
\]
In equilibrium we have that
\[ g_i^t = p'(x(\tau)) \text{ for all } i \in N \]
\[ \alpha + \beta y_i(\tau) = a - b \sum_{i=1}^{n} y_i(\tau) - \tau_i \text{ for all } i \in N \]
\[ \beta y_i(\tau) = a - \alpha - \tau_i - b \sum_{i=1}^{n} y_i(\tau) \text{ for all } i \in N \] (15)

Summing over \( i \) to \( n \) gives
\[ \beta \sum_{i=1}^{n} y_i(\tau) = n(a - \alpha) - \sum_{i=1}^{n} \tau_i - nb \sum_{i=1}^{n} y_i(\tau) \]
\[ (\beta + nb) \sum_{i=1}^{n} y_i(\tau) = n(a - \alpha) - \sum_{i=1}^{n} \tau_i \]
\[ \sum_{i=1}^{n} y_i(\tau) = \frac{n(a - \alpha)}{\beta + nb} - \frac{1}{\beta + nb} \sum_{i=1}^{n} \tau_i \] (16)

Inserting (16) in (15) this gives
\[ y_i(\tau) = \frac{1}{\beta} \left( a - \alpha - \tau_i - \frac{bn(a - \alpha)}{\beta + nb} + \frac{b}{\beta + nb} \sum_{i=1}^{n} \tau_i \right) \]
\[ = \frac{1}{\beta} \left( \frac{\beta(a - \alpha)}{\beta + nb} + \frac{b}{\beta + nb} \sum_{i=1}^{n} \tau_i - \tau_i \right) \] (17)

Let \( A = \frac{\beta(a - \alpha)}{\beta + nb} \), and \( B = \frac{b}{\beta + nb} \)
\[ y_i(\tau) = \frac{1}{\beta} \left( A + B \sum_{i=1}^{n} \tau_i - \tau_i \right) \] (18)

Using (18) the welfare function (2) reduces to
\[ W_i(\tau) = \frac{1}{\beta} \left( A \tau_i + B \sum_{i=1}^{n} \tau_i - \tau_i^2 \right) - c_i \] (19)

In order to calculate the toll-level of each country in equilibrium non-cooperative game theory is used. Let \( \langle N, T, W \rangle \) be a game where \( N = \{1, ..., n\} \) is the set of players (in this case the set of infrastructure managers), and for each player \( i \) a set of strategies \( T_i \). The set of strategies of a player \( i \) here consists of the set of possible toll levels. The set of all possible strategy profiles of the players is given by \( T = \times_{i \in N} T_i \). For each player \( i \) and strategy profile \( \tau \in T \) the function \( W_i(\tau) \) specifies the welfare level of player \( i \), in this model given by function (19).
In order to find the Nash equilibrium of the game corresponding to the parallel model, we need to calculate the best response function for every player \( i \), i.e. the the toll level that maximizes the welfare of country \( i \) given the toll levels in all other countries. The best response function of infrastructure manager \( i \) is given by the first-order condition\(^4\)

\[
\frac{dW_i(\tau)}{d\tau_i} = \frac{1}{\beta} \left( A + B \sum_{i=1}^{n} \tau_i + B\tau_i - 2\tau_i \right) = 0
\]

\[
(2 - B)\tau_i = A + B \sum_{i=1}^{n} \tau_i \quad \forall i \in N
\]  

(20)

Since this is true for all \( i \in N \) we can rewrite (20) as

\[
(2 - B)\tau_i = A + Bn\tau_i
\]

\[
\tau_i^* = \frac{A}{2 - Bn - B} \quad \text{for all } i \in N
\]  

(21)

The unique Nash equilibrium of the game is when every infrastructure manager set the toll \( \tau_i^* = \frac{A}{2 - Bn - B} \). Inserting (21) in the welfare function (19) gives the welfare level in equilibrium for non-cooperation.

\[
W_i^* = \frac{1}{\beta} \left( A - \frac{A}{2 - B(n + 1)} + Bn \frac{A}{2 - B(n + 1)} \right) \frac{A}{2 - B(n + 1)} - c_i
\]

\[
= \frac{A^2}{\beta (2 - Bn + B)^2} (2 - Bn + B - 1 + Bn) - c_i
\]

\[
= \frac{A^2}{\beta (2 - B(n + 1))^2} (1 - B) - c_i
\]  

(22)

If the infrastructure managers instead were to cooperate they would maximize the sum of welfare functions:

\[
\sum_{i=1}^{n} W_i(\tau) = \frac{1}{\beta} \left( A \sum_{i=1}^{n} \tau_i + B \sum_{i=1}^{n} \tau_i \sum_{j=1}^{n} \tau_j - \sum_{i=1}^{n} \tau_i^2 \right) - \sum_{i=1}^{n} c_i
\]  

(23)

with respect to \( \tau_i \) for all \( i \in N \). The first order condition\(^5\) then becomes

\[
\frac{\partial}{\partial \tau_i} \sum_{j=1}^{n} W_j = \frac{1}{\beta} \left( A + 2B \sum_{j=1}^{n} \tau_j - 2\tau_i \right) = 0 \quad \text{for all } i \in N
\]

---

\(^4\)Looking at the demand function (17) it is reasonable to limit the toll to the interval \( \tau_i \in [0, n(a - \alpha)] \). Since \( W_i \) is a continuous function and we optimize over a compact set, a maximum exist. Checking the second order condition shows that the first order condition gives a maximum.

\(^5\)Since the toll \( \tau_i \) is limited to the interval \( \tau_i \in [0, n(a - \alpha)] \) for all \( i \in N \), and \( \sum_{i=1}^{n} W_i \) is a continuous function and we optimize over a compact set there exist a maximum. Using the Kuhn-Tucker conditions shows that the first order condition gives a maximum.

19
Since this is the case for all \( i \in N \) we can rewrite this as
\[
A + 2Bn\tau_i - 2\tau_i = 0
\]
\[
\tau_{i}^{\text{coop}} = \frac{A}{2(1-Bn)}
\] (24)

Inserting (24) in the welfare function (23) we get
\[
W_{i}^{\text{coop}} = \frac{1}{\beta} \frac{A}{2-2Bn} \left( A - \frac{A}{2-2Bn} + Bn \frac{A}{2-2Bn} \right) - c_i
\]
\[
= \frac{1}{\beta} \frac{A^2}{(2-2Bn)^2} (2 - 2Bn - 1 + Bn) - c_i
\]
\[
= \frac{1}{\beta} \frac{A^2}{4(1-Bn)} - c_i
\] (25)

Comparing the welfare levels of cooperation (25) and non-cooperation (22) gives
\[
\frac{1}{\beta} \frac{A^2}{4(1-Bn)} - \sum_{i=1}^{n} c_i \geq \frac{A^2}{\beta (2 - B(n+1))^2} (1 - B) - \sum_{i=1}^{n} c_i
\]
\[
\frac{1}{4(1-Bn)} \geq \frac{(2 - B(n+1))^2}{(2Bn + B)^2} \geq 4(1-Bn)(1-B)
\]
\[
4 - 4Bn - 4B + 2B^2n + B^2 + B^2n^2 \geq 4 - 4Bn - 4B + 4B^2n
\]
\[
B^2(2n + 1 + n^2) \geq 4B^2n
\]
\[
2n + 1 + n^2 \geq 4n \text{ for all } n \geq 1, \text{ with equality for } n = 1
\]
showing that it is Pareto efficient for the infrastructure managers to cooperate. This is not surprising, the parallel paths are assumed to be perfect substitutes and it is therefore a competitive situation. Cooperation means that they can act as a monopoly. This implies that the welfare of countries outside the model, i.e. countries not owning links in the network, will be reduced by cooperation among the network owners.

In the non-cooperative situation one might expect the toll to equal marginal cost. From the welfare function
\[
W_i = \tau_i y_i(\tau) - c_i
\]
it is possible to write the first order condition for the non-cooperative situation as
\[
\tau_i = -y_i(\tau) / \frac{dy_i(\tau)}{d\tau_i}
\]
\[
= -y_i(\tau) / \left( \frac{1}{\beta} \left( \frac{b}{\beta + nb} - 1 \right) \right)
\]
\[
= y_i(\tau) / \left( \frac{1}{\beta} \left( 1 - \frac{b}{\beta + nb} \right) \right)
\]
\[
= \beta y_i(\tau) \frac{\beta + nb}{\beta + nb - b}
\]
where $\beta y_i(\tau)$ can be interpreted as the marginal external cost of congestion. Since

$$\frac{\beta + nb}{\beta + nb - b} > 1$$

the toll will exceed the marginal cost. The interpretation is that a high toll on one path increases congestion on the other paths, thus allowing tolls higher than marginal costs even in the case of competition (see for instance Verhoef et al. (1996) and van Dender (2005)). The resulting toll-level depends on the number of competing paths $n$. The larger the number of competing paths $n$ the closer to marginal cost pricing we get. This is reasonable since congestion decreases when $n$ increases.

Note also that

$$\tau_i^{\text{coop}} = \frac{A}{2(1 - Bn)} = \frac{\beta(a - \alpha)}{\beta + nb} = \frac{2}{2(1 - \frac{ln}{{\beta + nb}})} = \frac{\beta(a - \alpha)}{2(\beta + nb - nb)} = \frac{a - \alpha}{2}$$

which shows that the optimal toll in the cooperative case is independent of the congestion parameter $\beta$. This follows from the fact that the paths were assumed to be perfect substitutes, and the monopolistic pricing in the cooperative case.

### 6 Sharing profit - the parallel case

In order to simplify the analysis above we assumed $\alpha_i = \alpha$, $\beta_i = \beta$ for all $i \in N$. This resulted in a uniform toll. However, if the parameters are individual the toll will differ between the paths depending on the individual parameters.

**Example 6.1** Consider a case with three parallel paths and $a = 14$, $b = 1$, $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 1.2$, $\beta_1 = \beta_2 = 1$ and $\beta = 1.2$. This results in the tolls $\tau_1 = \tau_2 \approx 2.64$ and $\tau_3 \approx 1.14$.

In this section the more general case with individual parameters $\alpha_i$ and $\beta_i$ is considered. To model the parallel situation as a cooperative game $(N, v)$ we need to define a characteristic function $v$ assigning a value to every coalition $S \in 2^N$. However, just like in the serial case, the value of a coalition $S$ clearly depends on the toll levels of parallel links owned by players outside the coalition, i.e. links owned by players in $N \setminus S$. The standard way to handle this give rice to the same problems as in the serial case (see section 4), the players outside the coalition would have to behave in a way that is not credible,
and the value of each coalition would be \( v(S) = 0 \), with exception of \( v(N) \). Instead the characteristic function is here defined specifically for our model.

It is not credible that a coalition acts other than to maximize its welfare. However it is not always that players manage to cooperate even if this would be beneficial. Therefore let \( \tau_S \) be a vector of tolls for the countries in \( S \), and \( T(S) \) the set of all toll vectors for coalition \( S \). Let \( \tau_S^* \in T(S) \) be the profile of tolls that maximizes the welfare of coalition \( S \) given \( \tau_{N\setminus S} \), i.e.

\[
\tau_S^* \in \arg \max_{\tau_S \in T(S)} \sum_{i \in S} W_i(\tau_S, \tau_{N\setminus S})
\]

\[
= \left\{ t_S \in T(S) : \sum_{i \in S} W_i(t_S, \tau_{N\setminus S}) \geq \sum_{i \in S} W_i(\tau_S, \tau_{N\setminus S}) \text{ for all } \tau_S \in T(S) \right\}.
\]

Let \( d(S) \) be a coalition partition of the members of \( S \), consisting of a coalitions \( S_1, \ldots, S_k \), and \( D(S) \) the set of all possible coalition divisions of \( S \).

**Definition 6.2** A parallel transport network game \( (N, v) \) is defined as follows:

\[
N \subset \mathbb{N} \text{ is a set of players each owning one path of a parallel network.}
\]

\[
v(S) = \min_{d(N\setminus S) \in D(N\setminus S)} \sum_{i \in S} W_i(\tau_S^*, \tau_{(N\setminus S)_1}^*, \ldots, \tau_{(N\setminus S)_k}^*)
\]

i.e. the value of coalition \( S \) is the maximal welfare \( S \) gets when every coalition maximizes its welfare, and the players outside \( S \) divide into the coalition division that is least beneficial for coalition \( S \).

This definition is reasonable also when we do not limit the model to linear cost and demand functions.

In the serial case it seemed reasonable to allocate more to a country with high costs than to a country with low cost. Since the parallel case is a competitive situation, it is reasonable that the individual costs \( c_i \) are carried exclusively by the country itself. This turns out to be the case with the Shapley value for the class of parallel network games. Rewriting the welfare function as

\[
W_i(\tau) = f(\tau) - c_i
\]

where \( f(\tau) = \tau_iy_i(\tau) \) the welfare function can be rewritten as

\[
v(S) = \min_{d(N\setminus S) \in D(N\setminus S)} \sum_{i \in S} W_i(\tau_S^*, \tau_{(N\setminus S)_1}^*, \ldots, \tau_{(N\setminus S)_k}^*)
\]

\[
= \min_{d(N\setminus S) \in D(N\setminus S)} \sum_{i \in S} f_i(\tau_S^*, \tau_{(N\setminus S)_1}^*, \ldots, \tau_{(N\setminus S)_k}^*) - \sum_{i \in S} c_i
\]

\[
= \bar{v}(S) - \sum_{i \in S} c_i
\]
Using the definition of the Shapley value one can see that an individual constant does not change the allocation to other players.

\[
\Phi_i = \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} (v(S \cup \{i\}) - v(S)) \\
= \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} \left( \tilde{v}(S \cup \{i\}) - \sum_{j \in S} c_j - c_i - \left( \tilde{v}(S) - \sum_{j \in S} c_j \right) \right) \\
= \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} \left( \tilde{v}(S \cup \{i\}) - c_i - \tilde{v}(S) \right) \\
= \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} \left( \tilde{v}(S \cup \{i\}) - \tilde{v}(S) \right) - \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} c_i \tag{26}
\]

the number of coalitions with \( k \) members that can be formed from the set \( N \setminus \{i\} \) is \( \frac{(n-1)!}{k!(n-1-k)!} \) using this we can rewrite (26) as

\[
\Phi_i = \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} \left( \tilde{v}(S \cup \{i\}) - \tilde{v}(S) \right) - \sum_{k=1}^{n} \frac{(n-1)!}{k!(n-1-k)!} \frac{k!(n-1-k)!}{n!} c_i \\
= \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} \left( \tilde{v}(S \cup \{i\}) - \tilde{v}(S) \right) - \sum_{k=1}^{n} \frac{1}{n} c_i \\
= \sum_{S:i \notin S} \frac{|S|!(n-1-|S|)!}{n!} \left( \tilde{v}(S \cup \{i\}) - \tilde{v}(S) \right) - c_i
\]

which shows that every player will carry their own individual cost. This follows from the additivity property of the Shapley value.

Since the class of transport network problems defined in section 4 applies also for the parallel model, we can use the proportional rule \( PR \), the adjusted proportional rule \( APR \), and the adjusted equal profit rule \( AP \) also for the class of parallel transport network games. However, for these rules the individual cost for one country affects the allocation to another, which is not wanted in the parallel case.

An intuitive property of a solution concept is that it should allocate the profit in such a way that no player or coalition of players would have been better off by them selves. The set of such allocations is called the core.

**Definition 6.3** The core of the game \( \langle N, v \rangle \) is the set

\[
C(v) := \left\{ x \in \mathbb{R}^N \mid \sum_{i=1}^{n} x_i = v(N) \text{ and } \sum_{i \in S} x_i \geq v(S) \text{ for all } S \in 2^N \setminus \emptyset \right\}.
\]

It is obvious that we want a solution concept that yields an allocation in the core for every parallel network game. As it turns out, the Shapley value does not only yield an allocation in the core for every parallel network game, but also coincides with the bary-center of the core.
Definition 6.4 A game $(N, v)$ is convex if
\[ v(S) + v(T) \leq v(S \cup T) + v(S \cap T) \] for all $S, T \in 2^N \setminus \emptyset$.

Theorem 6.5 For a parallel network games $(N, v)$ the Shapley value is the bary-center of the core.

Proof. It is well known that the Shapley value of a convex game is the bary-center of the core of the game, see Shapley (1971). From the characteristic function it follows that
\[
 v(S) = \min_{d(N \setminus S) \in D(N \setminus S)} \sum_{i \in S} W_i \left( \tau^*_S, \tau^*_{(N \setminus S)_1}, \ldots, \tau^*_{(N \setminus S)_k} \right)
\]
\[
 \leq \min_{d(N \setminus (S \cup T)) \in D(N \setminus (S \cup T))} \sum_{i \in S} W_i \left( \tau^*_S, \tau^*_T, \tau^*_{(N \setminus (S \cup T))_1}, \ldots, \tau^*_{(N \setminus (S \cup T))_k} \right)
\]
\[
 v(T) = \min_{d(N \setminus T) \in D(N \setminus T)} \sum_{i \in T} W_i \left( \tau^*_T, \tau^*_{(N \setminus T)_1}, \ldots, \tau^*_{(N \setminus T)_k} \right)
\]
\[
 \leq \min_{d(N \setminus (S \cup T)) \in D(N \setminus (S \cup T))} \sum_{i \in T} W_i \left( \tau^*_T, \tau^*_S, \tau^*_{(N \setminus (S \cup T))_1}, \ldots, \tau^*_{(N \setminus (S \cup T))_k} \right)
\]
thus
\[
 v(S) + v(T) \leq \min_{d(N \setminus (S \cup T)) \in D(N \setminus (S \cup T))} \sum_{i \in T} W_i \left( \tau^*_S, \tau^*_T, \tau^*_{(N \setminus (S \cup T))_1}, \ldots, \tau^*_{(N \setminus (S \cup T))_k} \right)
\]
\[
 + \min_{d(N \setminus (S \cup T)) \in D(N \setminus (S \cup T))} \sum_{i \in S} W_i \left( \tau^*_S, \tau^*_T, \tau^*_{(N \setminus (S \cup T))_1}, \ldots, \tau^*_{(N \setminus (S \setminus T))_k} \right)
\]
\[
 \leq \min_{d(N \setminus (S \cup T)) \in D(N \setminus (S \cup T))} \sum_{i \in S \cup T} W_i \left( \tau^*_{S \cup T}, \tau^*_{(N \setminus (S \cup T))_1}, \ldots, \tau^*_{(N \setminus (S \cup T))_k} \right)
\]
\[
 \leq v(S \cup T)
\]
showing that a parallel network game is convex. Hence the Shapley value of a parallel network game is the bary-center of the core. \(\square\)

Moreover, Sprumont (1990) showed that the Shapley value of a convex game yields a population monotonic allocation scheme: the payoff allocated to each player according to the Shapley-value increases as he joins larger coalitions.

Due to all the nice properties of the Shapley value for the class of parallel network games, it is recommended as allocation rule for the parallel problem.
7 Local traffic

To simplify analysis in the previous sections demand for local traffic was assumed to be zero. Adding local traffic will of course affect both optimal toll and welfare levels. However, it will not change the recommendation for what allocation rule to use. In fact, the game theoretical analysis becomes almost identical. Adding local traffic means that the welfare function will be of the form \( W_i = f(t, \tau) - c_i \) for all \( i \in N \), which shows that both the serial and the parallel problems belongs to the class of transport network problems also when local traffic is added. As stated in section 5 the allocation rules \( PR \), \( APR \), and \( AP \) are efficient, individually rational, \( P \)-monotonic and \( W \)-monotonic for this whole class of problems. The \( PR \), and the \( AP \) rules are also self-consistent.

In section 6 it was shown that for the Shapley value an individual constant, such as the maintenance cost \( c_i \), does not affect the allocation to other players. This means that every player carries his/her own maintenance cost both in the serial and parallel case also with local traffic. Following the reasoning in section 5 and 6, this is reasonable in the parallel case but not wished for in the serial case.

The class of parallel transport games introduced in section 6 has a characteristic function defined for parallel problems without local traffic. However, it can easily be adapted to include local traffic by adding local tolls in the following manner; let \( \tau_S \) be a vector of tolls for transit traffic, \( t_S \) a vector of local tolls for the countries in \( S \), and \( \phi_S \) a vector of all tolls for the countries in \( S \), where the first \( |S| \) elements are the elements of vector \( \tau_S \) and the next \( |S| \) elements are the elements of \( t_S \). Let \( T(S) \) be the set of all such toll vectors for coalition \( S \). Let \( \phi_S^* \in T(S) \) be the profile of tolls that maximizes the welfare of coalition \( S \) given \( \phi_{N\setminus S} \), i.e.

\[
\phi_S^* \in \arg \max_{\phi_S \in T(S)} \sum_{i \in S} W_i(\phi_S, \phi_{N\setminus S})
\]

\[
= \left\{ \psi_S \in T(S) : \sum_{i \in S} W_i(\psi_S, \phi_{N\setminus S}) \geq \sum_{i \in S} W_i(\phi_S, \phi_{N\setminus S}) \text{ for all } \phi_S \in T(S) \right\}
\]

Let \( d(S) \) be a coalition division of the members of \( S \), consisting of the coalitions \( S_1, ..., S_k \), and \( D(S) \) the set of all possible coalition divisions of \( S \).

**Definition 7.1** A parallel transport network game \( \langle N, v \rangle \) is defined as follows:

\[
N \subset N \text{ is a set of players each owning one path of a parallel network.}
\]

\[
v(S) = \min_{d(N\setminus S) \in D(N\setminus S)} \sum_{i \in S} W_i(\phi_S^*, \phi_{(N\setminus S)_1}^*, ..., \phi_{(N\setminus S)_k}^*)
\]

i.e. the value of coalition \( S \) is the maximal welfare \( S \) gets when every coalition maximizes its welfare, and the players outside \( S \) divides into the coalition division that is least beneficial for coalition \( S \).

With this definition the analysis is analogous to section 6, i.e. the Shapley value is the bary-center of the core also for this class of games. The recommended allocation rules for the serial and parallel models are therefore identical with and without local traffic.
8 Conclusion

In this paper two types of transport network models are studied; one serial transport network (a transport corridor) and one parallel transport network where the parallel links are substitutes. For these models two types of analysis are performed. First the toll and welfare levels with and without cooperation are studied, using non-cooperative game theory. The analysis show that there are strong incentives for cooperative behavior among the countries owning links in the network. Without cooperation the parallel case is a competitive situation. It is therefore of no surprise that cooperation leads to higher tolls on transit traffic and higher welfare for the owners of the network. Countries outside the network will experience a reduced welfare due to the higher tolls for transit traffic in the network. It turns out that without cooperation the tolls slightly exceed marginal costs. The interpretation is that a high toll on one link increase congestion on the other links, thus allowing tolls higher than marginal cost even in the case of competition. It is shown that the toll converges towards marginal cost when the number of parallel links increase.

In the serial case it turns out that cooperation does not only increase the welfare of the owners of the serial network, but also the welfare in countries outside the network. There are a number of reasons for this. First of all, cooperation among the countries along the transport corridor will in fact reduce the tolls. Further without cooperation all decisions concerning maintenance and infrastructure investments are made on local level, and the paper shows that such decisions might be inefficient concerning the total welfare level. Therefore, without regulation or cooperation the tolls will be higher than what is efficient, while the standard of infrastructure will be lower than what is efficient.

For cooperation to occur it is not enough that the total welfare level increases compared to non-cooperation, the countries also have to be able to agree on how to split the resources raised from cooperation. The analysis shows that this cannot be done via a uniform toll level and each country keeping their own toll incomes. In the parallel case this is obvious since in equilibrium the total user cost, including tolls, must be equal for every link, and since other user costs vary, so must the tolls. In the serial case it is unreasonable to expect a country with a very costly link to accept setting the same toll as a country with a less costly link.

Instead of setting a uniform toll, the total income from cooperation has to be allocated among the cooperating countries. Of course no "correct" such allocation exist; however, some allocations are more likely to be accepted than others. These are allocations that satisfy intuitive properties related to fairness. By supplying such rules, negotiation costs are reduced and cooperation more likely to occur.

The second type of analysis performed in the paper deals with such allocations for the serial and parallel transport models. To be able to analyze the cooperative situation thoroughly, cooperative game theory was used. For this purpose a new class of problems is introduced - transport network problems. Both the serial and the parallel model fits in this class of problems. Further a new class of cooperative games is introduced - the class of parallel transport network games. This class of games corresponds to problems like the
parallel model.

The Shapley value is one of the most well-known solution concepts in cooperative game theory. In the parallel case it is easy to motivate the use of the Shapley value since it has very nice properties for the class of parallel transport network games, such as being the bary-center of the core of the game. In the serial case however, the Shapley value coincides with setting a uniform toll level and letting each country keep their toll incomes. As mentioned above this is not a reasonable allocation. Instead, three new allocation rules are introduced; the proportional rule \((PR)\), the adjusted proportional rule \((APR)\) and the adjusted equal profit rule \((AP)\). These rules allocate more to a country with large costs than to a country with low costs, which is reasonable for the serial case. They also have a number of other nice properties.

In most of the analysis the demand for local traffic is assumed to be zero. Adding local traffic will of course affect both optimal toll and welfare levels. However, the game theoretical analysis becomes almost identical when adding local traffic, and it does not effect the properties of the studied allocation rules. The recommendations for cooperative solutions are therefore identical with and without local traffic. However, although intuitive, the paper does not prove that it is beneficial to cooperate when there is a demand for local traffic.

References


Appendix E. Sharing costs in Swedish road ownership associations
Sofia Grahn-Voorneveld
Sharing costs in Swedish road ownership associations

Sofia Grahn-Voorneveld*

Swedish National Road and Transport Institute
P.O.Box 55685, 102 15 Stockholm, Sweden

Abstract

In Sweden a large part of the road system is privately owned. Most of these privately owned roads are rural roads used by farmers and summer cottage owners, or used for forest transport. The roads are mainly provided by ownership associations.

The motivation of this paper is the practical problem of how such an ownership association can divide the costs for the road network among the members in a “fair” way. The problem is treated from a game theoretical point of view, making use of the Shapley value. This means that the problem is associated with a game - a mathematical representation of the conflict situation. The Shapley value is a very important solution concept for cooperative games, like the game in this case. For games corresponding to this specific type of problems, it is shown that the Shapley value has excellent properties, such as being an element of the core, and being very easy to compute.

*I thank Jan-Eric Nilsson for useful comments.
1 Introduction

Sweden is a sparsely populated country, with large areas that are unpopulated or extremely sparsely populated. For this reason a community does not have to provide roads in areas outside the city plan. Therefore it is not surprising that a large part of the Swedish road system is privately owned. Most of these privately owned roads are rural roads used by farmers and summer cottage owners, or used for forest transport. The roads are mainly provided by ownership associations. These ownership associations are formed in cooperation with the National land survey of Sweden - Lantmäteriet, and the rules for membership and many other aspects are regulated by law. Any real property, within the area for the road system of the association, can be forced to join the association if the benefit from the road system is considered to be essential for the real property in question. In connection with the constitution of the road association, it is decided what share of the costs, for building and maintaining the road, each real property has to carry. The share is related to characteristics of the property, not to the actual use of the road. An obvious characteristic is the road length needed to connect a property to the public road network. Other possible characteristics are whether the real property is a summer cottage, all year around house or a farm. If the property is a forest area the share is usually related to the size of the forest area. Today the shares are decided by the road association or the National land survey of Sweden. If the road association decides the shares, they have to be approved by the National land survey to make sure that they are "reasonable". However, the National land survey does not have rigid rules for what the shares should be, only guidelines.

Many of the road associations get contributions from both the government and communities. However, to get such contributions the roads have to be open to the public. The fact that the membership of a road association normally is not voluntary and that non-members cannot be excluded from using the roads, means that a road ownership association cannot be viewed as a club\textsuperscript{1}.

The motivation of this paper is the practical problem of how a road association can divide the costs among the members in a "fair" way. The costs are of two different types. The first type are costs that are directly related to the usage of the road, i.e. the marginal cost. Even though the actual use is not measured it is possible to let every user pay roughly their own cost. However, a large part of the costs relate to building costs or maintenance costs related to weather rather than usage, for instance road damage in connection to thawing of the ground. The problem of how to divide this second type of costs is treated from a game theoretical point of view, making use of the Shapley value. For this purpose a new class of problems are defined - the class of road association problems. Each such problem can be associated with a cooperative game - a mathematical representation of the conflict situation.

Even though fairness is a subjective property and there exist no "correct" cost allocation to the problem some allocations are more likely to be acceptable than others. These

\textsuperscript{1}In club theory the membership of a club should be voluntary and it must be possible to exclude non-members from using the club facility.
are allocations that satisfy properties related to fairness. The Shapley value is a very important solution concept for cooperative games since it satisfies many such properties. For games corresponding to road association problems it has excellent properties such as being an element of the core. Further, the Shapley value can be written as a very simple expression for this type of game, which makes it very easy to calculate.

2 The model

The road system of a road association can have very different structures. To create a model to handle this some definitions from graph theory\(^2\) are used.

**Definition 2.1** A graph is a pair \((V, L)\) where \(V\) is a non-empty set of elements called nodes (or vertices or points), and \(L\) is a finite set of unordered pairs of distinct elements of \(V\) called links (or edges).

\(^2\)In mathematics, graph theory is the study of mathematical structures used to model pairwise relations between objects from a certain collection.
Definition 2.2 A *root* is a node that connects the road association network to the public road network or other facility of importance to the members of the road association, such as a harbour.

It is possible that the graph corresponding to a road association network has several roots.

The problem for the road association is to divide the costs of its road network. The costs are of two different types. The first type of cost is directly related to how much a road is used. Since the actual use is not measured, the real properties are classified depending on their expected usage. The set \( N = \{1, ..., n\} \) of members of the road association is divided into \( k \) groups with \( n_1, ..., n_k \) members in each group respectively. Each group \( g_i \) is associated with a maintenance cost \( \alpha_i \), which is the expected maintenance cost caused by a member of this group. Since this type of cost is independent of the needs or deeds of other actors, it is reasonable that every real estate carries its own costs, i.e. real estate \( i \) pay \( \alpha_i \).

The second type of costs are fixed with respect to the usage. Instead they are related to the different needs for sophistication of the road network. An obvious sophistication factor is the road length. Example of other sophistication factors are bearing capacity and winter standard. To be able to handle the second type of costs a new class of problems is defined.

**Definition 2.3** A *road association problem* is a tuple \((N, T, F, f, c)\) consisting of

- a finite set of members \( N \subset \mathbb{N} \);
- a graph \( T \) consisting of nodes, roots and a set \( L \) of links.
- a set \( F \subset N \) of sophistication factors;
- a function \( f : N \to 2^{L \times F} \) specifying for every member which links he needs with what sophistication factor;
- a function \( c : L \times F \to \mathbb{R}_+ \) specifying the cost of each link with respect to every sophistication factor.

The set \( 2^{L \times F} \) denotes the set of all possible combinations of links and sophistication levels. A player who does not need a specific link use this link with a sophistication level associated with a cost equal to zero.
Example 2.4 Consider a road association with the following graph:

\[ \begin{array}{c}
\text{root} \\
\text{A} \\
\text{B} \\
\text{C} \\
1 \\
2 \\
3 \\
4 \\
5 \\
\end{array} \]

Figure 3: The graph

The graph has three links A, B and C. Link A is needed by all five real estates of the road association, although real estate 1 needs the link with a lower sophistication level than the rest of the real estates. Link B is needed by real estates 2 and 3. Real estates i.e. 1, 4 and 5 do not need this link, and therefore use the link with a sophistication level associated with a cost equal to zero. Link C is needed by real estates 4 and 5, but not by real estates 1, 2 and 3 which therefore use it with a sophistication level associated with a cost equal to zero.

It is possible to order the users/real properties in terms of their needs with respect to every sophistication factor. Accommodating a user with a certain need allows accommodating all users with lower needs at no extra cost. This type of problem is known as an airport problem, since it originally was phrased as the problem of financing a landing strip. Every airport problem can be associated with a coalitional form game called an airport game.

A cooperative cost game is a tuple \( (N, c) \) where \( N = \{1, ..., n\} \) is the set of players. The set of all possible coalitions of players in \( N \) is denoted by \( 2^N = \{ S \mid S \subseteq N \} \). The function \( c : 2^N \rightarrow \mathbb{R} \) is called the characteristic cost function of the game, and assigns to each coalition \( S \) a cost \( c(S) \in \mathbb{R} \), with \( c(\emptyset) = 0 \).

Definition 2.5 Airport game: Let \( N \) denote the set of players. The set \( N \) is split into \( k \) groups of players \( g_1, ..., g_k \) with \( n_1, ..., n_k \) players in each group respectively. Each group is assigned a non-negative number \( b_1, ..., b_k \). The airport game corresponding to \( g_1, ..., g_k \) and \( b_1, ..., b_k \) is the cooperative cost game \( (N, c) \), where \( N \) is the set of players \( N = \bigcup_{i=1}^{k} g_i \), and the cost function \( c \) is defined by

\[
c(S) = b_1 + ... + b_{m(S)}
\]

for every \( S \subseteq N \), where \( m(S) = \max \{ m : S \cap g_m \neq \emptyset \} \).
An airport game is a cost game for the financing of one facility where the needs of the coalitions are linearly ordered. The groups \( g_1, \ldots, g_k \) are groups of players with different desires for sophistication levels (originally length of landing strip). Every \( b_l \) represents the extra building cost needed to adjust a facility, used by the less sophisticated players of groups \( g_1, \ldots, g_{l-1} \), to be used also by the more sophisticated players in group \( g_l \).

For a road association problem every link of the road network corresponds to an airport game for every sophistication factor. Let \( G_{ij} \) be the airport game corresponding to link \( i \) and sophistication factor \( j \in F \). The members of the road association are divided into \( k \cdot (ij) \) groups of players \( g_{ij}^1, \ldots, g_{ij}^{k(ij)} \) with \( n_{ij}^1, \ldots, n_{ij}^{k(ij)} \) players in each group respectively. Every group \( g_{ij}^q \) is associated with a cost \( b_{ij}^q \) which is the cost to raise the standard of the road from what is needed by group \( g_{ij}^q \) to the standard needed by group \( g_{ij}^{q-1} \). Members of the road association who do not need this link has sophistication level zero associated with a cost equal to zero.

**Definition 2.6 Road association game:** Every road association problem corresponds to a game \( G = \langle N, c \rangle \) which is the sum of a finite set of airport games, where each link corresponds to an airport game for every sophistication factor. The cost function \( c \) of the game \( G \) is defined by

\[
c(S) = \sum_{i \in L, S \cap N_i \neq \emptyset} \sum_{j \in F} \left( b_{ij}^1 + \ldots + b_{ij}^{m_{ij}(S)} \right)
\]

for every \( S \subseteq N \), where \( m_{ij}(S) = \max \{ m : S \cap g_{ij}^m \neq \emptyset \} \); i.e. \( g_{m_{ij}(S)} \) is the group with the highest sophistication level containing members of \( S \).

This means that the \( c(S) \) is the cost needed to build/maintain the road network with a sophistication level, with respect to all sophistication factors, so that all members of \( S \) can use the road network.

**Example 2.7** Consider the graph in example 2.9. Assume that the real properties 2, 3, 4 and 5 are summer cottages while real property 1 is a farm with need for higher bearing capacity. The problem has two sophistication factors; the road itself (factor 1) and bearing capacity (factor 2). Each link therefore corresponds to two different airport games, one for each sophistication factor. For every game the players are divided into different groups depending on their need for sophistication level. Every such group is associated with a cost as described above. The costs for this example are given in the table below.

The game \( A_1 \) relates to the costs for building and/or maintaining link \( A \) with a normal standard. The members of the road segment either use part of the link (player 1) or use the whole link (players 2, 3, 4, 5). These two groups therefore corresponds to different sophistication levels, with respect to road length. The groups are denoted by \( g_{A_{11}}^1 = \{1\} \) and \( g_{A_{11}}^2 = \{2, 3, 4, 5\} \). The set \( g_{A_{11}}^1(\text{Player 1}) \) is of a lower sophistication level than the rest of the players, who need the whole road segment of link \( A \). The cost \( b_{A_{11}}^1 \) is the cost for building/maintaining the lowest sophistication level, i.e. the road section used by player...
1. The cost $b_2^{A1}$ is the cost for raising the road standard to sophistication level 2, i.e. the costs for the rest of the road segment.

The game $A_2$ relates to the costs for sophistication factor 2 - bearing capacity. Player 2, 3, 4 and 5 need no extra bearing capacity than what is included in normal road standard covered by sophistication factor 1. The cost associated with their sophistication level is therefore zero. Hence, $g_1^{A2} = \{2, 3, 4, 5\}$ is associated with the cost $b_1^{A2} = 0$. Player 1 has a need for a higher sophistication level. Thus $g_2^{A2} = \{1\}$ is associated with the cost $b_2^{A2}$ which is the cost to maintain the road section of link A with the extra bearing capacity needed by player 1.

The group division and costs for the games corresponding to links $B$ and $C$ are shown in the table below:

<table>
<thead>
<tr>
<th>Game</th>
<th>$b_1^{ij}$</th>
<th>$b_2^{ij}$</th>
<th>$g_1^{ij}$</th>
<th>$g_2^{ij}$</th>
<th>$n_1^{ij}$</th>
<th>$n_2^{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>1</td>
<td>2</td>
<td>${1}$</td>
<td>${2, 3, 4, 5}$</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>$A_2$</td>
<td>0</td>
<td>2</td>
<td>${2, 3, 4, 5}$</td>
<td>${1}$</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>$B_1$</td>
<td>0</td>
<td>4</td>
<td>${1, 4, 5}$</td>
<td>${2, 3}$</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$B_2$</td>
<td>0</td>
<td>–</td>
<td>${1, 2, 3, 4, 5}$</td>
<td>–</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>$C_1$</td>
<td>0</td>
<td>2</td>
<td>${1, 2, 3}$</td>
<td>${4, 5}$</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$C_2$</td>
<td>0</td>
<td>–</td>
<td>${1, 2, 3, 4, 5}$</td>
<td>–</td>
<td>5</td>
<td>–</td>
</tr>
</tbody>
</table>

where $n_{ij}^{ij}$ is the number of members in group $g_{ij}^{ij}$. The road association game $G$, corresponding to this road association problem, is the sum of the above 6 airport games.

3 Dividing the costs

The problem that every road association face is how to divide the costs for its road network in a "fair" way among the members. A road association game $G$, corresponding to a road association problem, describes the costs that each coalition would face if the coalition were to act separate from the players outside this coalition. Now an allocation rule allocating the costs among the players has to be chosen.

A very important solution concept for transferable utility games is the Shapley value, which has excellent properties and has been applied successfully in cost allocation problems (see Shapley (1953), Tijs and Driessen (1986), Young (1994), Moulin and Schenker (1996) and Fragnelli et al. (1999)).

The Shapley value was introduced by Shapley (1953) and can be characterized by four properties. Somewhat informally:

- **efficiency**, i.e. no resources are wasted;
- **anonymity**, i.e. two identical players are treated equally;
- **dummy property**, i.e. a player with a constant marginal contribution to every coalition of which he is a member, is allocated this constant.
• **additivity**, i.e. the solution of the sum of two games, is the sum of the solution to the two games.

The Shapley value is the only solution concept satisfying all four of these axioms.

In this context the Shapley value assigns to each player the average marginal cost contribution that the player makes to each of the coalitions to which he belongs.

Let \( \Phi_i(c) \) denote the Shapley value of a cost game \( \langle N, c \rangle \), then the Shapley value can be written as

\[
\Phi_i(c) = \sum_{S \in \mathcal{F}} \frac{|S|!(n-1-|S|)!}{n!} \left( v(S \cup \{i\}) - v(S) \right) \quad \text{for player } i \in N
\]

In a practical environment it is extremely important that the solution allocation can be computed easily. For an airport game the Shapley value can be written as

\[
\Phi_i(c) = \sum_{l=1}^{q(i)} \frac{b_l}{\sum_{p=l}^{n_p} n_p} \quad \text{for all } i \in N
\]

(see Littlechild and Owen (1973)). For a road association game, the Shapley value can be written as the sum of the Shapley values of the link games of the corresponding road association problem. This follows from the additivity property of the Shapley value. Thus:

\[
\Phi_k(c) = \sum_{i \in L, k \in N_i} \sum_{j \in F} \sum_{l=1}^{q(i)} \frac{b_{l}^{ij}}{\sum_{p=l}^{n_p} n_p} \quad \text{for all } k \in N
\]

**Example 3.1** Using the same road association as in example 2.7 the corresponding road association game \( G \) is the sum of 9 airport games. Using expression (1) we get

<table>
<thead>
<tr>
<th>Game</th>
<th>player 1</th>
<th>player 2</th>
<th>player 3</th>
<th>player 4</th>
<th>player 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{5} + \frac{2}{4} = \frac{13}{20} )</td>
<td>( \frac{1}{5} + \frac{2}{4} = \frac{13}{20} )</td>
<td>( \frac{1}{5} + \frac{2}{4} = \frac{13}{20} )</td>
<td>( \frac{1}{5} + \frac{2}{4} = \frac{13}{20} )</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( B_1 )</td>
<td>0</td>
<td>( \frac{3}{5} = 2 )</td>
<td>( \frac{3}{5} = 2 )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( B_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( C_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{1}{2} = 1 )</td>
<td>( \frac{1}{2} = 1 )</td>
</tr>
<tr>
<td>( C_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>( \frac{13}{5} )</td>
<td>( \frac{27}{10} )</td>
<td>( \frac{27}{10} )</td>
<td>( \frac{17}{10} )</td>
<td>( \frac{17}{10} )</td>
</tr>
</tbody>
</table>

resulting in the Shapley value \( \left( \frac{13}{5}, \frac{27}{10}, \frac{27}{10}, \frac{17}{10}, \frac{17}{10} \right) \).

An intuitive property of a solution concept is that it should allocate the profit in such a way that no player or coalition of players would have been better off by them selves. The set of such allocations is called the core.

**Definition 3.2** The core of the cost game \( \langle N, c \rangle \) is the set
\[ C(c) := \left\{ x \in \mathbb{R}^N \mid \sum_{i=1}^{n} x_i = c(N) \text{ and } \sum_{i \in S} x_i \leq c(S) \text{ for all } S \subseteq 2^N \setminus \emptyset \right\}. \]

It is well known that the Shapley value of an airport game lies in the core of the game. See for instance Thomson (2007). Due to the additivity of the Shapley value also the Shapley value of a road association game lies in the core.

### 4 Conclusion

In this paper the problem of dividing costs for a road association network over its members is addressed. The costs are of two different types. The first type are costs that are directly related to the usage of the road, i.e. the marginal cost. Even though the actual use is not measured it is possible to let every user pay roughly their own costs. However, a large part of the costs are not related to the actual usage. The problem of how to divide this second type of costs is treated from a game theoretical point of view. The problem has a structure where it is possible to order the users in terms of their needs, with respect to every sophistication factor. Accommodating a user with a certain need allows accommodating all users with lower needs at no extra cost. This type of problem is known as an airport problem. Every airport problem can be associated with a coalitional form game called an airport game. Therefore the problem of dividing the second type of costs over the members of the road association is modelled as a game, which is a sum of a finite number of airport games.

The Shapley value is one of the most important solution concepts for cooperative games. The Shapley value of a cost game assigns to each player the marginal cost contribution that the player makes to each of the coalitions to which he belongs. The marginal cost contribution in this context is the cost the player adds to the coalition if he has a higher need for sophistication then the rest of the players of the coalition, i.e. the sophistication has to increase so that also this player can use the road network. Apart from the nice properties that defines the Shapley value, the Shapley value always lies in the core of a game corresponding to a road association problem. Further, the Shapley value for games corresponding to road association problems can be written as a simple expression. This makes it very easy to calculate the Shapley value for this type of problem. The Shapley value therefore is a very appealing solution concept for road association problems.

### References:


