E C O N O M I C R E S E A R C H C E N T R E

REPORT OF THE
HUNDRED AND SEVENTH ROUND TABLE
ON TRANSPORT ECONOMICS

held in Paris on 26th-27th march 1998
on the following topic:

USER CHARGES FOR RAILWAY INFRASTRUCTURE

EUROPEAN CONFERENCE OF MINISTERS OF TRANSPORT

TABLE OF CONTENTS

GERMANY

Gerd ABERLE
Professor of Transport Economics
University of Giessen
Germany

SUMMARY

Giessen, November 1997

1. WHY ARE USER CHARGES FOR RAILWAY INFRASTRUCTURE NECESSARY?

Until the end of the 80s, there were only state-owned railway undertakings in western Europe, and they owned both the railway networks and the rolling stock (= vertically integrated railway). Other railway undertakings were operated only as works or regional railways with a feeder function. The state-owned railways had a monopoly of supply in transregional passenger and freight rail transport.

As from the beginning of the 60s, there was a continual and serious deterioration in the economic situation of all state railways. The reasons were, in particular:

- Strong and sometimes cut-throat competition from road haulage;
- The dominant influence of politics on the strategic and operational decisions of the railway undertakings;
- Inadequate technical and economic adaptation to market needs;
- Inadequate capital endowment by the state owner, resulting in an economically insupportable level of indebtedness (despite very high annual state subsidies to the railways).

The thorough-going structural reforms of the railways implemented since the beginning of the 90s in several western European countries are intended to:

- Enable rail transport to achieve economic equilibrium and hence ease the financial burden on the state budget;
- Create the preconditions for maintaining or increasing the rail transport market share.

Significant support for this was given by the Railway Directive 91/440 issued by the EC Council of Ministers in 1991, which were supposed to be implemented in the EC Member States by 1 January 1993 (though this was not the case even by the end of 1997 in all EU countries). Directive EG 95/19 of 19 July 1995 is important for the levying of charges for track use.

One of the main components of the structural reforms was (and is) the separation of rail infrastructure from operations, i.e. vertical disintegration. This separation means that a distinction is to be made between track or network companies on the one hand, and passenger and freight train operators on the other.

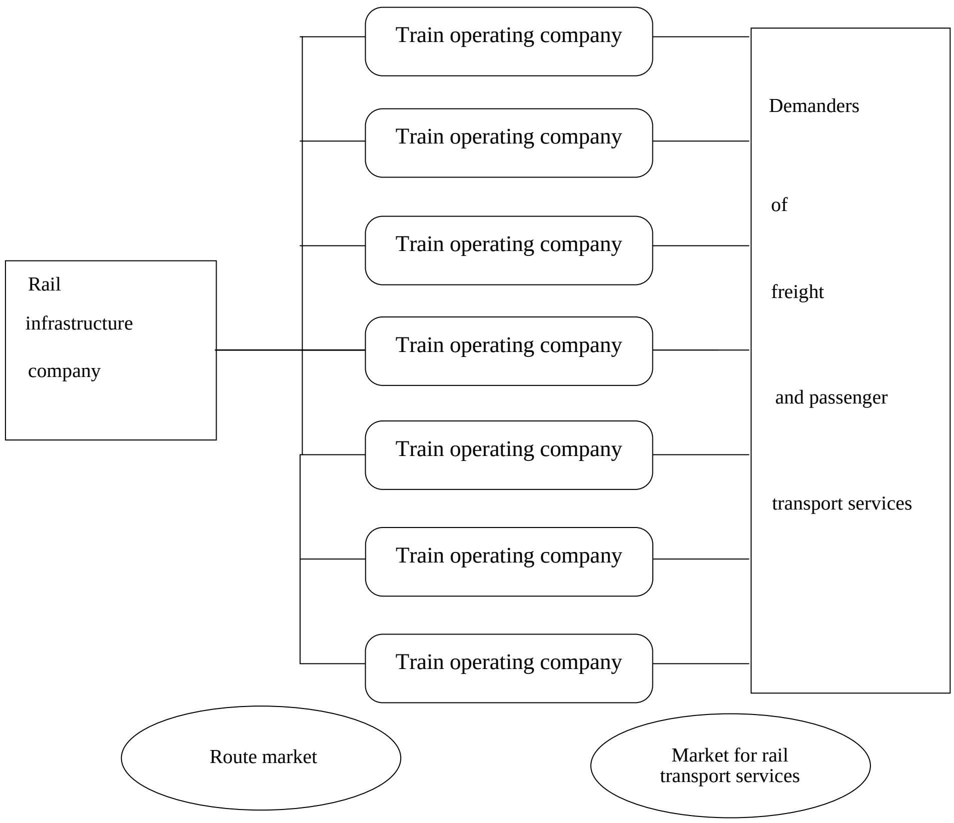
Although EC Directive 91/440 required only separate accounts and gave a positive opinion on organisational separation, certain countries have already implemented institutional (actual) separation (e.g. Sweden, the United Kingdom) or have taken preparatory steps towards this (e.g. Germany).

Largely independent of the question of the separation of track and operating companies is that of the privatisation of state railways. In the vertical disintegration process, a distinction must be made in this connection between the privatisation of the train operating companies and that of the railway infrastructure companies. This is also of importance for the discussion of charges for the use of rail infrastructure.

The main reason for the need to develop special pricing for the use of the rail infrastructure lies in the opening of the network to other railway undertakings. This possibility of permitting competition in the field of rail transport, opened up only to a very limited extent by EC Directive 91/440, has already been implemented in a very much more liberal fashion in some countries.

This opening of the networks to third parties is forcing the railways to do a lot of rethinking. Whereas intermodal competition, in particular with the roads and to some extent with the waterways and air transport (market competition), was often a difficult challenge for the former vertically-integrated railways, a new market dimension for railways is now appearing. With the separation of track and train operations, routes now have to be marketed. The supplier is the institution with ownership rights to the network (the railway infrastructure company); the demand comes from train operating companies, the number of which is increasing with the opening of the network to third parties. In addition to the so far mainly state-owned operating companies that have emerged from the earlier vertically integrated railways, foreign and private national operating companies are now appearing as demanders of routes. The new market situation brought about by vertical disintegration and the opening of networks to third parties is shown in Figure 1.

Figure 1. **Structure of the rail transport market**



The market for the right to use railway infrastructure becomes all the more significant:

– the more intensive the use of the network by third-party domestic train operators, and
– the stronger the demand for infrastructure services by foreign railway companies.

The proposal made by an international group of experts ("Wise men's report", June 1996) included in the Commission of the European Union White Paper (Com 96/421, July 1996), *A strategy for revitalising the Communities'* *railways*, to establish Cargo Freeways in EU Member countries, also includes the requirement to develop pricing systems for the use of the corresponding rail routes.

In some European countries, no charge is made for the use of the track, even though at least an accounting and organisational separation of infrastructure from operations has been implemented. This temporary or durable renunciation of charging for the use of parts of the railway infrastructure is not economically justified. It is more a case of political decisions which distort competition. The reasons for not charging for track use are, in particular:

- The intermodal competition faced by train operators, in particular with road and waterway transport, does not permit the imputation of track costs;
- There is no institutional (actual) separation of infrastructure from operations yet, but only an accounting and organisational separation. Opening the network to other rail transport companies is not envisaged, or only in a few exceptional cases, so that there is no necessity for a pricing system for the use of railway infrastructure;
- The design and practical implementation of a pricing system for railway infrastructure use is extraordinarily difficult. In the case of mixed networks, used for local and mainline passenger traffic as well as freight traffic, the track costs are mainly common costs resulting from shared use of the track. The imputation of track costs to the individual track user in the sense of “true costs” is not possible.

These three main arguments against charging for the use of railway infrastructure are not convincing however:

1. The railway track represents a very substantial capital investment. A general renunciation of any imputation of the costs of using this resource precludes any possibility of economic mechanisms for controlling the use of part of the nation’s productive capital. The associated subsidising of train operators sends out false price signals to the demanders of transport services. Since the economic worth of a rail route is determined by the characteristics of its use (for example: time frame, different quality requirements of high-speed and regional passenger trains and freight trains), this should also be reflected in user charges.

2. The argument that imputing the network costs to operators is not possible because of intermodal competition, so that a pricing system does not make sense, simply does not stand up. This calls into question the marketability of railway transport services.

   Only if the competing modes did not have to pay for the use of their transport infrastructure might the taking over (by the State) of the rail infrastructure costs be envisaged. However, such global financing of infrastructure by the taxpayer offends against the basic economic principle already mentioned under (1) above, and therefore cannot be justified.

   Furthermore, because of the differing infrastructure costs of the competing transport modes, if the State were to take over all infrastructure costs it would be subsidising transport services to differing extents, and hence again sending out false signals for the modal split.

   Problems arise, however, when differing levels of infrastructure cost coverage are demanded for the intermodal competitors. In practice, this applies in particular to the very low infrastructure cost coverage in waterway transport and parts of the road haulage sector. In such situations, it should be considered whether the financial goals of the infrastructure supplier should be formulated accordingly. The Directives for the overall degree of cost coverage to be achieved in the case of railway infrastructure could be set somewhere below 100 per cent to correspond with the degree of coverage of the intermodal competitors.

3. The mention of the difficulties involved in preparing and implementing a pricing system for track use is, in principle, correct but much theoretical preparatory work has been done even for the complex situation of mixed networks. Chapter 3 goes into these studies in detail. Since this is largely new territory for transport policy and there is very little empirical experience, further research and discussion is required on this question. Round Table 107 can make a significant contribution here.

2. REQUIREMENTS OF A PRICING SYSTEM FOR THE USE OF RAILWAY INFRASTRUCTURE

2.1. Micro- and macroeconomic aims of pricing systems

On the market for railway infrastructure services, the supply and demand appears at route level. As a rule, it is a case of a network supplier which was part of the unified railway undertaking that existed before vertical disintegration. Here, there is a monopoly supplier of railway infrastructure. On the demand side are the passenger and freight train operating companies, made up of the state railway companies and, in the case of the opening of the network to third parties, other domestic and foreign train operating companies. In this case, there is intramodal competition (competition in the field).

The microeconomic goal of the supplier of network services is then to fully or partly cover the costs of maintaining and operating the rail network by means of user charges and appropriate marketing. Here another question arises: that of whether full cost coverage, or possibly even a profit, is at all realistic. This question is not discussed in this paper. The financial goals may, in any event, be very different depending on the political goals, the intensity of the intermodal competition and the absolute level of network costs.

On the microeconomic level, the network supplier must be given the possibility of differentiating the rail infrastructure use charges (the route prices) according to specific criteria. Details are given in Chapter 3. This means that there is no uniform route price per reference unit (e.g. train-kilometre), but rather prices which depend on the cost and demand conditions (price elasticities, specific requirements for the routes used, time frames, etc.). An important feature of price formation must be non-discrimination. The danger of discrimination is always relatively great when there are legal interconnections between the network company and certain train operating companies. This is the case, when:

- there is no complete vertical disintegration of the railway undertaking, but only accounting and organisational separation of infrastructure from operations; or
- the legally independent infrastructure and train operators are run and co-ordinated by a single holding company.

In both cases, the interests of the enterprise as a whole are dominant. The competition of third parties on the network, i.e. their demand for network slots as a rule means competition for the train operating companies of the group. Here, preventing market access for third parties on the network, or making it difficult, leads to protection for these operating companies. The prices charged to third parties can play an important role if they differ from those asked of the train operators associated with the network company without there being any special requirements to justify it.

In concrete terms: there is price discrimination if the operating companies of the group are charged lower route prices for otherwise equal conditions than third-party train operating companies. This can be done through:

- the use of unpublished settlement prices for organisationally associated companies or
- subsequent refunds on the route prices to the associated train operators by the infrastructure company.

In addition to price policy discrimination against third-party train operating companies, the (route) allocation practices of network companies have a high discrimination potential. This is not discussed any further here.

The macroeconomic goals of pricing systems for infrastructure use are:

- to make the user of railway infrastructure cover the cost of using resources; and
- to improve the market position of rail transport in general and not only of the already established train-operating companies. Through giving access to the network to third parties, intramodal competition is expected to promote keener pricing, better quality and more innovation.

2.2. Reference parameters for an infrastructure use pricing system

There are three aspects to be taken into account here.

- First, there is the service to be priced. The factors taken into account are:

    - Train operation over a section of route;
    - Coach or axle-kilometres over a section of route;
    - Train-kilometres over a section of route.

The parts of the network or sections of route differ in their capital and operating costs and in their economic worth (such as: capacity utilisation, quality, geographical location, etc.). It is necessary to break down the network into sections of route (links between nodal points). The use of the section of route may be best expressed in terms of *train-kilometres*. As a basis for charges, this unit promotes:

- Loading trains to full capacity with a given number of wagons;
- Formation of the longest possible trains.

Both effects enhance the efficiency of utilisation of the section.

– Second, charges are to be made for the use of stations (passenger stations, freight depots, marshalling yards) and sidings. These facilities must also be accessible for third-party train operators without discrimination, since without this possibility of use there can be no intramodal competition in the rail sector.

– Third, the cost categories relevant for price setting and the method of determining them on the basis of the services used by the rail infrastructure user have to be decided. This requires, in the first place, an analysis of the problem, taking into account the alternative procedures available according to price theory. This is the subject of the following chapter.

3. RAILWAY INFRASTRUCTURE PRICING

3.1. General introduction

A pricing system for the use of railway infrastructure has to satisfy several conditions.

– As already mentioned, it must be non-discriminatory, i.e. the network supplier must charge all demanders of network services the same price for the same things. Here there should be the greatest possible transparency of all conditions for present and potential users of railway infrastructure. Special settlement prices in the case of only accounting or organisationally, but not actually (institutionally) separated track and train operating companies, which differ from the published prices for otherwise identical conditions, are inadmissible. This also violates Article 8 of EC Directive 91/440 and possibly also Article 6.1 of this Directive.

– It should help to increase the capacity utilisation of railway infrastructure, and hence to make rail transport more attractive and increase rail transport service output;

– The pricing system should also help network companies to achieve their financial goals. These financial goals can been seen as:

  - Total cost coverage, including the notional rate of return on the fixed capital;
  - Total cost coverage excluding the notional rate of return on the fixed capital in the case of state investment decisions (Feldstein Thesis; M.S. Feldstein, 1964);
  - Coverage of a percentage of the total costs; here the "state share" (public interest) can be taken into account;
  - Coverage of the total network costs including external social costs resulting from the construction and maintenance of railway infrastructure (e.g. separation effects, land use). Noise and pollutant emissions are social (external) costs to be imputed not to the infrastructure, but to train operation.

– The construction of a pricing system for the use of railway infrastructure must also take account of which parameters are measurable and usable in practice. There are considerable differences between price theory modelling and practical implementation in the real world. They stem in particular from problems with the availability of data on the required parameters. The transaction costs arising in each case are also of importance.

3.2. Analysis of alternative pricing systems

The train operating companies express demand for timetable slots. A timetable slot is the right to run a specific train (weight, length, speed) over a specific route (between two nodes) at a specific time (day of the week, time). Because of the generally limited number of companies requiring slots, it is a case of demand oligopoly. The train operator's willingness to pay is determined in particular by the intermodal competition (road, waterway and, in some cases, air transport).

If we consider a given stretch of the rail network with only one rail infrastructure supplier, this is generally agreed to be a case of natural monopoly. It is defined by the subadditivity of the cost function.

In considering the short-term costs, i.e. within a given network, a striking feature is the very low marginal costs (or rather, incremental costs). As direct use-dependent costs, they only amount to 3-8 per cent of the total costs. They depend on train weights and speeds and on the condition of the track. Total average costs fall (economies of density).

Various route pricing systems can be used:

a) Monopoly prices

Assessment: In the case of both monopolistic unit prices and monopolistic price differentiation, the macroeconomic goal of achieving the greatest possible intensity of use of the track is lacking.

b) Average cost prices

Assessment: The advantage of simple calculation is offset by the disadvantage that no special account is taken of train operators who are prepared to bear the incremental costs directly associated with use. There is thus less than optimal allocation of network resources. Furthermore, the costs relevant to decisionmaking are incorrectly reflected by average cost calculation.

c) Incremental cost prices (in the sense of empirically determinable marginal costs)

Assessment: Total average costs cannot be covered by means of pure incremental cost pricing. The macroeconomically optimal pricing rule leads to microeconomically undesirable deficits.

The deficit problem can be eliminated, however, if a system of additional charges on top of the incremental costs is introduced. This can be achieved through single or multi-part pricing.

c.1)In a single-part pricing system, based on incremental costs with the condition that total costs must be covered, an additional charge is incorporated into the price. This additional charge to cover the costs which are not dependent on use may be linear, degressive or progressive for the user ("fully distributed costs") or be based on the operator's willingness to pay.

Additional charges according to willingness to pay take account of the different willingness and ability to pay of the different train operators. These charges can be differentiated according to the type of traffic: suburban passenger, mainline passenger, freight traffic. Furthermore, a difference can be made according to the time slot, the route and punctuality requirements. The least loss of macroeconomic welfare, as compared with pure incremental cost pricing, is achieved by using Ramsey prices.

c.2)A two- or multi-part pricing system shows separately the additional charges to the incremental costs. These additional charges can take account of:

-- Particular capacity shortage situations on heavily-trafficked stretches and at certain times (peak load pricing);

-- Departures from stretch-specific normal speeds (defined according to optimal capacity utilisation). These departures are relevant for trains with both higher and lower speeds, because faster trains too have a negative effect on capacity utilisation;
-- A basic contribution to cover the deficit arising with incremental cost pricing. This additional charge (also

known as the system charge) can be a fixed time-dependent contribution or can be graded according to demand characteristics (e.g. high-speed traffic, regional traffic, combined traffic, etc.). Differentiation according to spatial parts of the network is also possible.

The system charge represents an option price and works like a quantity rebate, since it does not depend on the number of slots demanded per unit of time (year, for example) or part of the network (and may be related to the whole network). This encourages the use of the network. The extent to which a differentiation of the system charge is actually made will depend on the associated transaction costs and its practicality.

The relatively most suitable pricing rule for railway slots is the use of multi-part tariffs, which on top of the incremental costs of each train passage have an additional system charge and, if appropriate, further additional charges in the case of peak periods and interference to traffic flow caused by speed differentials.

Since train operating companies with relatively low demand for slots are greatly penalised by the fixed system charge, there should be a possibility of choosing between different price structures. This is also necessary in order to avoid discrimination against train operators with a relatively weak market position.

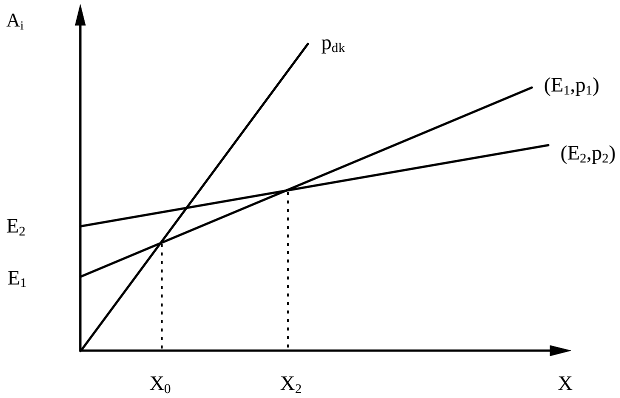
The possibility of choice is given by having alternative tariffs with different levels of system charges and correspondingly different use-dependent prices. Relatively low system charges are then associated with higher use-dependent price components. The train operator can then decide on the tariff which gives it the lowest route costs.

Figure 2. **Differentiated system charges with the possibility of choice**

Let the expenditure function of a train operating company demanding slots be

$A_i = f(E_i, p_i)$ where :

p_i is the price per slot, p_{dk} the average cost price, p_{ik} the incremental cost price
E_i is the system charge and
A_i is the total route expenditure for train operator $_i$.



A_i = expenditure of operator $_i$
E_1, E_2 = differentiated system charges
pi = variable price
$p_{dk} > p_1 > p_2 > p_{ik}$
x = planned quantity of slots demanded by operator $_i$

With a very low level of demand x_0 (in train-kilometres/year, for example) train operating companies may choose system charge E_1 and relatively higher use-dependent price *(p_1);* with higher demand x_2 , on the other hand, a higher system charge E_2 with relatively low use-dependent price components *(p_2)* may be chosen. The average cost price is represented by a constant increase with growing demand for slots *(p_{dk}).*

3.3. The problem of infrastructure cost responsibility

Rail infrastructure costs are very much influenced by the technical characteristics of the track (high-speed stretches, for example). As a rule, however, stretches are also used by trains that use only part of their technical possibilities. This applies above all to mixed traffic stretches. The question therefore arises of which trains have to bear which costs.

There are two concepts here:

– According to the *prime user concept*, trains, or train types, are ranked according to the technical and organisational demands on the infrastructure. The trains with the highest demands (high-speed trains, for example) bear all the total costs of the stretch except for the incremental costs caused by other users.

- According to the *sole user concept,* there is again a hierarchy, but in this case the type of train which is highest in the hierarchy pays, in addition to the incremental costs caused by it, only the direct, use-independent costs (fixed costs, overheads) which it actually needs for its supply. Unlike under the prime user concept, the sole user does not have to bear cost components which arise through investment, but which are not actually required by users. This increases the cost responsibility of the network company.

The prime user concept is the basis for route pricing in Japan (prime users are the passenger services; freight traffic pays only its incremental network costs) and in the USA (AMTRAK pays only the incremental costs for the use of freight traffic stretches). The sole user concept is used in the United Kingdom (sole users are the individual passenger transport companies).

The central problem with the use of both concepts lies in establishing the hierarchy. The sole/prime user may be, for example:

- passenger traffic, or
- freight traffic.

Further differentiation is both possible and necessary. Thus high-performance, fast traffic requires the most sophisticated track equipment and very often also expensive engineering structures such as tunnels and bridges. It is also possible to divide the network into subnetworks according to function: for example, a fast passenger traffic network, regional passenger traffic networks and a freight traffic network. A hierarchy can also be established within these functional subnetworks.

3.4. Efficiency assessment

The pricing systems described above, with additional charges on top of the incremental costs, do not automatically lead to pressure to improve productive efficiency (favourable cost situation). The network company endeavours to cover its total costs through the additional charges, which means that a profit in the sense of a rate of return on capital is possible. Falling costs can be reflected in falling route prices; rising costs result in an increase in the additional charges.

Incentives to increase productive efficiency, on the other hand, always exist if the network company is organised as a profit centre and the profits realised do not have to be paid out to an associate undertaking or a holding company. It can be deduced from this that true incentives to reduce costs exist only if the network company is privatised and there is an appropriate pricing system. This appropriate pricing system from the standpoint of productive efficiency is that of multi-part tariffs. It is particularly well-suited in the case of progressive additional charges on top of incremental costs, where the progressive additional charges (the higher the demand for the slots, the higher the additional charge) are particularly effective in promoting competition and hence have a cost-reducing effect. With such a pricing system, new demanders appearing on the market (private regional train operators, for example) with relatively low additional charges receive the same consideration as state railways with a strong market position. Also positive is the differentiation of the additional charges according to willingness to pay, as this gives the possibility of setting the slot prices flexibly in the case of low willingness to pay. It would thus be possible, in the case of low willingness to pay and very underutilised sections of route, to reduce the price to the level of the incremental costs or to impute only a small proportion of the system charge. In this way, not only can rail traffic be retained or even won, but also there can be some relief of sections of the rail network where the capacity limits have be reached or where there is already congestion.

On the other hand, relatively high system charges and special peak charges and, where appropriate, speed-dependent charges (in the sense of opportunity costs) if train speeds disturb operations, can be levied if certain stretches or parts of the network have very high capacity utilisation with congestion effects. Here, the nodal points (above all stations) in particular are to be taken into account. The node performance is much more difficult to improve than stretch performance. In the latter case, "driving by electronic sight" by means of Computer Integrated Railroading (CIR) can bring significant capacity increases (up to 30 per cent) even on already existing stretches, while in the case of nodal points such improvement requires substantial additional construction investment.

3.5. The problem of the duration of track-use contracts

Basically, the market for railway routes is no spot market. Instead there must, as a general rule, be relatively long contracts. Both the suppliers of routes (the infrastructure companies) and the train operating companies have a direct interest in this.

Of considerable importance for the length of track use contracts are the transaction costs, in particular, the following components:

- Factor specificity or irreversibility;
- Transaction frequency;
- Uncertainty.

Factor specificity or irreversibilities are of considerable importance for the route market, since special track equipment (superstructure, electronic train control, signalling, etc.) is frequently necessary only for certain potential users. On the other hand, the train operators are very interested in it so as to be able to prepare timetables for relatively long periods and offer specific time slots for trains. A reduced transaction frequency reduces total transaction costs. Longer contract periods also reduce the uncertainties for actors on the route market.

Longer-term contracts can also have negative effects on the efficiency of the route market, however, insofar as they cause:

- obstacles to market access for third parties; or
- irreversibilities with sunk cost effects for train operating companies.

Market obstacles for third parties can appear in the case of routes which are heavily-trafficked and of strategic importance for timetables. In the case of long-term contracts, these can be used very little if at all by intramodal competitors. The purchase of route use rights to block intramodal competition is to be seen as abuse of market power with highly discriminatory effects.

The following rules can be applied to reduce the risks resulting from long-term contracts:

- For the network operator, price escalator clauses can protect against significant cost increases or sharp increases in demand. The train operator’s contract would then ensure route availability, but not the right to a fixed route price for the whole period;
- For the train operator, the possibility of trading the right to use the route would make it possible to withdraw from a contract little or no sunk cost losses. The risk of loss is not eliminated by this, but only reduced.

3.6. Some implementation problems for a route pricing system

The introduction of a route pricing system for railway infrastructure involves finding solutions to a number of problems. Some important aspects are summarised in what follows.

– Determining the incremental costs for each train category and section of route;
– Determining the stretch-specific optimal train speeds in order to fix additional charges for higher and lower speeds;
– Delimitation of congested stretches in order to fix peak charges;
– Taking into account of what is known as system traffics, which regularly (often according to regular-interval timetables) occupy routes for relatively long periods of time;
– Delimitation of the rail installations to be attributed to the track;
– Development of a pricing system for the use of stations, marshalling yards and sidings;
– Establishing the differentiation in the system charges (e.g. passenger traffic/freight traffic; high-performance *versus* regional traffic);
– Establishing demand groups to estimate the willingness to pay;
– Price policy treatment of empty trains.

Considerable progress has been made towards solving the problem of delimiting the railway equipment to be attributed to the track. Here we refer to Regulation (EEC) 2598/70 of 18 December 1970 Annex 1, Part A, and Regulation (EEC) 1108/70 of 4 June 1970, Annex 1. According to the latter, the railway infrastructure consists of: ground area, track and track bed; engineering structures (bridges, tunnels, etc.); passenger and goods platforms; level crossings; superstructure, accesses for passengers and goods, including access by road; safety, signalling and telecommunications installations; lighting installations; plant for transforming and carrying electric power for train haulage; buildings used by the infrastructure department, including installations for the collection of transport charges.

4. THE DEUTSCHE BAHN AG (DB AG) ROUTE PRICING SYSTEM

4.1. Introduction

The comprehensive structural reform of the railways in Germany was prepared, to a large extent, through the concrete recommendations of an independent government commission. This commission of eleven members was set up by the Federal Government in September 1989 and presented its final report in December 1991. In addition to the proposals for cancelling the debt, conversion to a joint-stock company and restructuring the staff, central considerations were the organisational and legal separation of railway infrastructure from train operations and the substantial opening of the network to third parties.

Already on 1 January 1994, after the necessary amendment of certain articles of the German constitution and a considerable number of specialised Acts, the structural reform of the railways was implemented. The legal opening of the network to other train operating companies (in addition to the traffic divisions of DB AG) made it necessary to set up a pricing system for the use of railway infrastructure. This was a difficult task, since:

- the DB AG network, with some 42 000 km of track, was run mainly as a mixed network for both passenger and freight traffic, with very different speeds and axle loadings;
- there was no model for the preparation of a pricing system suitable for a mixed use complex, so that new ground had to be broken.

Another factor to be taken into account is the tight deadline; the first version of a pricing system for the use of the DB AG network was published as early as 1 August 1994. In 1995, there was a minor revision, concerned above all with the level of the quantity discounts. In the first version of the route pricing system quantity discounts (based on train-kilometres) up to a maximum of 20 per cent were provided for in both passenger and freight traffic. This led to discrimination against third-party train operators, however, since in practical terms only DB AG was able to take advantage of these discounts. Criticism of this original arrangement led to modifications and a maximum discount of 5 per cent.

To complement this route pricing system, DB AG produced price lists for the use of stations and sidings. The price to be paid for the use of railway infrastructure is therefore generally made up of three components:

- The route price as payment for the use of sections of route;
- The station price as payment for the use of stations;
- Payment for the use of sidings.

What follows is concerned solely with route prices.

4.2. Structure of the route price system

The annual costs of the DB AG infrastructure amount to between 9-11 billion DM; precise figures are not available, because the company does not present network cost and earnings information in its annual reports. Furthermore, the legal provisions for the reform of the railways contain no stipulations concerning whether the prices paid for the use of routes are to aim at full cost coverage, partial cost coverage or a profit.

The DB AG route price system is based on the pricing of basic routes between nodes. The unit for price setting is the train-kilometre. For a train working over several nodes, the corresponding route prices are added together to give the total price for the entire route. A quantity or time discount may be deducted from this total, in order to obtain the actual price to be paid for the use of the railway infrastructure.

The route price is subject to the legal value-added tax. The differentiations between route prices result from the two parameters determining the prices of the sections of route concerned:

- Characteristics of the section of route, expressed in the *route category;*
- Characteristics of the train working over this section of route, expressed in the *train price class*.

The DB network is divided into ten route categories. Each route category reflects the network function of the section concerned in the sense of its transport importance and the quality (possible high-speed operation).

The train price classes take account of the wear and tear effects on the track and the superstructure and the required timetable quality (tolerance for departures from the desired journey time). Both characteristics of the train price class (wear and tear, timetable quality) are taken into account through weighting factors. The basic price (which depends on the route category) is multiplied by these in order to obtain the route price. These prices are published in the DB AG route price book for each section of route. There are seven train price classes in passenger and five classes in freight traffic.

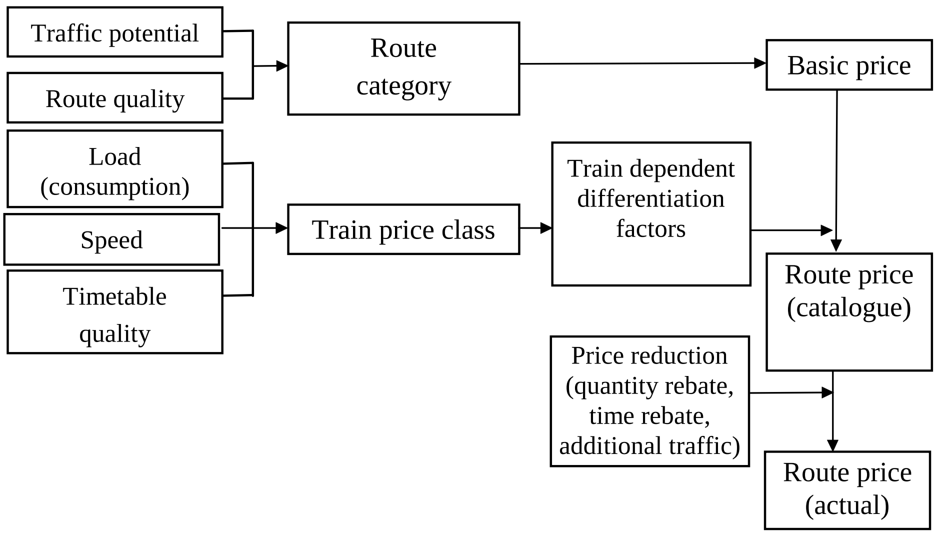
The weighting factors for wear and tear lie between 0.9 for light and 1.1 for very heavy trains. The different timetable quality requirements have weighting factors between 0.8 and 1.2.

Each of the basic routes (links between nodes) appearing in the DB route price book thus takes into account:

- the importance of the stretch in the overall network;
- the operating speed;
- the train weight and speed, and
- the desired timetable quality.

Figure 3 illustrates these parameters.

Figure 3. **Route price determinants**



The Annex to this report presents examples of route prices for passenger and freight traffic over selected transport links. Examples are given for high-speed and regional passenger trains and for various types of freight train.

The quantity discounts that can be obtained are intended to reflect the quantity-dependent cost savings. They have different graduations for passenger and freight traffic. In passenger traffic, the quantity discounts begin at 18 million train-kilometres/year, in freight traffic at 15 million train-kilometres/year (0.5 per cent discount in each case). The maximum discount rate of 5 per cent is reached at 180 million train-kilometres/year in mainline passenger traffic and 150 million train-kilometres/year in freight traffic. In suburban passenger traffic, there is a special discount which begins already at 100 000 train-kilometres/year (0.5 per cent) and rises to a maximum of 5 per cent.

The time discount for longer-term route contracts begins at two years (2 per cent) and rises with contract periods from five years to a maximum of 6 per cent.

Before taking the discounts into account, the average route prices per train-kilometre in suburban passenger traffic are between DM 8.54 and DM 10.43; in mainline passenger traffic between DM 18.75 and 22.49, and in freight traffic between DM 11.12 and DM 15.28.

In 1995, a special discount system for suburban passenger traffic was introduced. The explanation for this is as follows.

In Germany, a regionalisation of local public passenger traffic was introduced in connection with the structural reform of the railways. Local rail transport services were contracted out by regional authorities to DB AG (local passenger traffic division) or to other train operators. Insofar as these authorities demand additional local rail passenger services and routes as compared with the past, the train-kilometre price is only DM 5.

4.3. Remarks on the DB AG route price system

This route price system for a European railway, produced in a relatively short time, certainly takes into account important cost determinants, but is only to be seen as a first step towards an efficient rail infrastructure pricing system.

Important problem areas are:

1) The present DB AG route price system is a case of an average cost system modified by various influencing parameters (total route costs, divided by the total train-kilometres operated by the railway).

2) The route price system is based on a single-part tariff. The necessary differentiations in route prices cannot be sufficiently achieved with this pricing system. The system is inflexible with respect to peak period loadings of stretches on certain days of the week and at certain times.

3) As a result, this means that the prices for low-traffic stretches are too high and those for heavily-trafficked and congested stretches too low. The promotion of rail transport in intermodal competition requires, for example, prices approaching the (very low) incremental costs for the use of low-utilisation stretches. Conversely, on high-utilisation congested stretches, the opportunity costs should be taken into account; they require higher usage prices than provided for in the present route price system.

4) The route price system applies for all demanders of rail infrastructure services. Furthermore, it is not possible to check whether the traffic divisions of DB AG have to pay the same route prices (taking quantity and time discounts into account) in all cases as other train operators. It at least cannot be excluded that the train operators associated with the network company pay more favourable prices than the published route prices. Herein lies a discrimination potential to the disadvantage of third parties.

5) There is a further discrimination problem in local rail passenger traffic insofar as all of the routes operated in Germany qualify for the quantity discount, regardless of the regional distribution. Local train operating companies other than DB AG are generally active only in certain regions and therefore have a relatively low number of train-kilometres. It is questionable whether the cost savings relevant for the quantity discount can be demonstrated even if all the train-kilometres of all the regional operators were added together. Local rail passenger transport always has a regional dimension in both the service provision and costs arising. In Germany, there is a draft Railway Infrastructure Use Regulation (EIBV). It is expected that this statutory instrument will come into force at the beginning of 1998. Regarding route prices (user charges), § 6 states that they can be made up of general basic prices for all types of transport or special basic

prices for certain types of transport plus or minus supplements and discounts. In the calculation of the additional charges, the route category, time slot, journey time and infrastructure wear and tear are to be taken into account. This formulation points to a split tariff.

In § 8 of this Regulation, quantity discounts are specifically defined as *stretch-related* reductions. They must not exceed the demonstrable cost savings.

6) Discrimination can arise not only through route pricing, but also through the way in which routes are allocated. So long as there is no complete institutional separation of network and train operator, the traffic divisions of the railway undertaking dominant on the market will try to keep competitors out by means of route availability. This can be done by the infrastructure division passing on information about the routes sought by third-party operators to the traffic division. The latter can then claim the same spatial and temporal requirements or buy route usage rights to block others.

   The potential combination of price and route availability discrimination substantially reduces allocative efficiency in the rail transport sector.

7) A substantial competition problem also arises from the fact that different countries deal with the route pricing problem in different ways. If in certain European countries there are expressly no route prices calculated, this leads to allocative inefficiencies. Not only is the domestic train operator made artificially cheaper by being subsidised, but intramodal competition between alternative routes is distorted to the disadvantage of those train operating companies that have to pay user charges. Such a situation is developing in European north-south traffic, since the Netherlands and now also Sweden do not impose user charges. Furthermore, the competition at seaports is distorted by artificially changed hinterland transport costs, since, for example, DB AG has to pay route prices for Hamburg, Bremen and

other North Sea and Baltic port hinterland transport, but Netherlands Railways (NS) do not have to pay user charges for Rotterdam and Amsterdam hinterland traffic.

8) Intermodal competition means that the need to impute user charges leads to the necessity to reduce infrastructure costs. This activates the efficiency potential that helps strengthen the position of rail transport. The high network costs of all railways represent a system problem for rail transport, all the more so as they can amount to up to 40 per cent of the total costs of rail transport. It is therefore necessary not only to make the network costs transparent, but also to create a pressure to durably reduce the network costs, through rationalisation investments, in order to be able to stand up to the intermodal competition. A route pricing system is a particularly well-suited indicator here.

   Lastly, it should be pointed out that agreement on a unified method of establishing user charges for European railway infrastructure is an important requirement. The still very different approaches to the calculation of user charges by the different European railways distort competition. From this standpoint all efforts to reach a scientific consensus on the method of calculating route prices are of considerable value for railway policy.

4.4. The new DB AG route price system in 1998

DB AG's first route price system, dating from 1994, will be replaced by a new system around mid-1998. The main reasons for this replacement are the problem areas previously mentioned in section 4.3. Further reasons are to be seen in the experience acquired so far with the present route price system, as well as additional organisational or institutional changes at DB AG.

On 1 January 1999, DB AG will thus be converted into a holding company (the new DB AG) responsible for the following management companies: DB Personennahverkehr AG (local passenger traffic), DB Personenfernverkehr und Touristik AG (long-distance passenger traffic), DB Cargo AG (freight traffic), DB Netz AG (network system) and DB Bahnhöfe AG (station system). DB Netz AG must, in the future, therefore publish route system revenues in the annual accounts. These revenues are from the route prices paid by the DB AG's three transport undertakings and by other route users.

Since 1994, important streamlining measures have made it possible to reduce network costs to about DM 7 billion a year. The current requirement in Germany to cover total network costs by route prices will therefore be somewhat easier to meet.

The new route price system is structured according to various parameters:

-- It defines six line categories into which the entire DB AG network is subdivided (K1 to K6). The categories reflect the varying technical and economic quality standards (maximum speeds, location of the lines classified, importance of the lines in terms of traffic). K1 is the most and K6 the least attractive. The prices are graded accordingly;
-- Three traffic categories are defined (long-distance passenger traffic, local passenger traffic and freight traffic). The route prices differ in each case and also take into account the relative ability to pay in addition to costs (prime user principle);
-- Apart from the line categories and types of traffic, the line load factor is also included in each network category. There are three load factor levels (BI to BIII);
-- The new route price system is designed as a two-part tariff system. The route prices to be paid by the rail transport undertakings consist of two basic amounts:
  - A price for the purchase of a *network card* for network categories K1 to K6 for the particular type of traffic. On each available network, minimum line lengths are applicable. The network card price is based on the number of *line* kilometres. The card permits the use of the network in the network category purchased and on the lines contained within it. The network card price is therefore an exercise price for use of the network;
  - In addition to the user right, obtained at a flat rate for a year by purchasing the network card, a second transport price component must also be paid for the *train* kilometres actually performed. Since the network card has already been purchased, these *variable* price components are set very low compared with the earlier DB AG route price system. The intention is thus also to create an incentive to perform more train kilometres and thereby make the most of the sliding scale effect for network card costs.
-- In addition to this two-part tariff, each route user who performs only relatively few train kilometres a year, must pay a so-called *variable charge*. This is a through charge which depends only on the *train kilometres*, the respective *network category* and the *load factor* of the lines on which these train kilometres are performed. In the case of the

variable charge, the train kilometre prices in each network category are considerably higher than the variable price component in the split charge applicable to a network card;

-- Network card purchasers can obtain a *time discount* if they undertake to acquire a card every year for a period of ten years. This time discount reflects the savings on transaction costs as well as the greater investment planning security for Netz AG.

Quantity discounts are not planned. The split charge, however, works as an implicit quantity discount for applications for a large number of routes.

The new route price system must comply with the non-discrimination principle, which is laid down in Section 15 of the General Railway Act (AEG), as well as in the Railway Infrastructure User Rules which came into force in Germany on 1 January 1998. In addition, consistency with the EU Council Directive 95/19 of 19 June 1995 (allocation of railway infrastructure capacity and the charging of infrastructure fees) is to be assumed.

The observance of this principle is particularly important, since in the second stage of German rail reform there is still no full legal separation between the DB AG's network system and transport undertakings. On the contrary, Netz AG will also be strategically positioned by Management Holding DB AG, as Netz AG's nominal capital will be fully held by DB AG.

5. SUMMARY

1) The deregulation of European railway markets has led to the legal separation and, in several cases, to the organisational separation of track and train operators. In some countries, institutional separation has already been completed or is being prepared.

2) One of the basic ideas of this vertical disintegration is to enhance the performance of rail transport by opening the networks to third-party train operators (intramodal competition). This requires a non-discriminatory charging system for infrastructure use. The creation of Cargo Freeways also calls for such a system.

3) User charges have a twofold task. On the one hand, they are required to promote the macroeconomically optimal use of the resources tied up in railway infrastructure; on the other, they are supposed to make it possible to have the user cover the network costs. Given the intensive intermodal competition on transport markets, there are potential contradictions here.

4) Relatively favourable preconditions for fulfilling these two objectives are given by a charging system in the form of a multi-part tariff based on train-kilometres. The basis is formed by the incremental costs of a train working over a given route. This is completed by additional charges which may depend on the train category (passenger/freight, special requirements regarding the technical characteristics of the track, train control, engineering structures, etc.), the degree of capacity utilisation of the stretch (peak load pricing), the willingness to pay of the demanders and the opportunity costs arising because of departures from the optimum operating speeds specific to the stretch.

5) The route pricing system introduced in Germany in 1994 is to be seen as a modified average cost pricing system. Important parameters for the route prices are the train axle loadings, train speeds, punctuality requirements and the importance of the stretch in the operation of the network as a whole. Since it is a single-part tariff, the system is as yet unsatisfactory from the standpoint of market-oriented price flexibility. A reform of the present route pricing system is envisaged for 1998.

6) In order to pursue an efficient European railway policy, agreement on a common method of determining user charges is required. In view of the still outstanding theoretical problems and the specific complexities of implementation, this requirement will be difficult to fulfil.

ANNEX

EXAMPLES OF DEUTSCHE BAHN AG
ROUTE PRICE CALCULATIONS

Example 1: Passenger traffic

Maschen-Munich-Pasing

Train price class: P1, High-speed traffic
Load: Maximum 1 000 t
Speed: 200 kph and over
Planning quality: Maximum: 105 per cent
Train category: ICE and similar

Train price class: P3, Mainline express traffic with regional connections
Load: Maximum 550 t
Speed: Up to 160 kph
Planning quality: Maximum: 110 per cent
Train category: Inter Regio and similar

Route section	Section Number	Route price in DM	
Hamburg hbf		Inter Regio	ICE
	54	148	177
Hamburg-Harburg			
	56	627	752
Lüneburg			
	68	578	693
Uelzen			
	75	1 551	1 861
Hanover Hbf			
	109	2 189	2 627
Göttingen			
	136	963	1 156
Kassel-Wilhelmshöhe			
	184	1 970	2 364
Fulda			
	250	2 014	2 417
Würzburg			
	269	1 371	1 645
Nürnberg Hbf			
	294	825	990
Treuchtlingen			
	311	1 121	1 345
Augsburg			
	338	908	1 089
Munich-Pasing			
		14 265	17 116

Example 2: Freight traffic

Maschen-Munich

Train price class: G1, Fast and high-value freight traffic
Load: Maximum 1 500 t
Speed: 120 kph and over
Planning quality: Maximum: 125 per cent
Train category: Express freight trains

Route section	Section Number	Route price in DM
Maschen		
	23	476
Lüneburg		
	42	2 174
Hanover (SFS)		
	121	2 090
Göttingen		
	153	235
Eichenberg		
	164	706
Bebra		
	189	670
Fulda		
	205	1 923
Würzburg		
	258	1 047
Ansbach		
	286	600
Treuchtlingen		
	293	659
Ingolstadt		
	303	932
Munich		
	313	118
Munich-East		
		11 630

According to the DEGT, Maschen-Munich East is 761 km.

Cost per train-kilometre: DM 11 630/761 = DM 15.28.

Example 3: Passenger traffic

Giessen-Frankfurt South

Train price class: P5, Regional passenger traffic
Load: Maximum 400 t
Planning quality: Maximum: 120 per cent
Train category: Regional express, semi-fast and similar

Route section	Section Number	Route price in DM
Giessen		
	46	282
Bad Nauheim		
	58	40
Friedberg		
	72	181
Bad Vilbel		
	83	161
Frankfurt Hbf		
	92	45
Frankfurt South		
		709

According to the DEGT, Giessen-Frankfurt South is 68 km.

Cost per train-kilometre: DM 709/68 = DM 10.43.

Example 4: Passenger traffic

Karlsruhe-Bretten

Train price class: P6, Local passenger traffic
Load: Maximum 400 t
Planning quality: Maximum: 120 per cent
Train category: Regional, CityBahn, local trains

There are no differences between P5 and P6 concerning the pricing criteria load and planning quality. Only the train categories differ.

Route section	Section Number	Route price in DM
Karlsruhe		
	76	49
Karlsruhe-Durlach		
	73	20
Grötzingen		
	74	136
Bretten		
		205

According to the DEGT, Karlsruhe-Bretten is 24 km.

Cost per train-kilometre: DM 205/24 = DM 8.54.

Example 5: Regional freight traffic

Giessen-Hanau Hbf

Train price class: G3, Heavy freight traffic
Load: Maximum 2 500 t
Speed: Up to 100 kph
Planning quality: Maximum: 150 per cent
Train category: Freight trains handed over complete or in wagon strings by the shipper and accepted by the consignee without any marshalling operations

Route section	Section Number	Route price in DM
Giessen		
	91	203
Butzbach		
	95	158
Friedberg		
	106	55
Assenheim		
	116	241
Hanau Nord		
	153	55
Hanau Hbf		
		712

According to the DEGT, Giessen-Hanau Hbf is 64 km.

Cost per train-kilometre:

G3: DM 712/64 = DM 11.13
G2: DM 704/64 = DM 11.00
G4: DM 690/64 = DM 10.78.

Example 6: Regional freight traffic

Giessen-Frankfurt South

Train price class: G3, Heavy freight traffic
Load: Maximum 2 500 t
Speed: Up to 100 kph
Planning quality: Maximum: 150 per cent
Train category: Freight trains handed over complete or in wagon strings by the shipper and accepted by the consignee without any marshalling operations

Route section	Section Number	Route price in DM
Giessen		
	91	203
Butzbach		
	95	158
Friedberg		
	115	203
Bad Vilbel		
	133	158
Frankfurt hbf		
	148	58
Frankfurt South		
		780

According to the DEGT, Giessen-Frankfurt Süd is 68 km.

Cost per train-kilometre:

G3: DM 780/68 = DM 11.47
G2: DM 704/68 = DM 11.34
G4: DM 690/68 = DM 11.12.

BIBLIOGRAPHY

Aberle, G. and A. Brenner (1994), "Trassenpreissystem für die Deutsche Bahn AG -- eine erste kartellrechtliche und ökonomische Beurteilung", *Internationales Verkehrswesen*, 46, pp. 704-712.

Aberle, G., A. Brenner, A. Hedderich (1995), "Trassenmärkte und Netzzugang", Band 8 der *Gießener Studien zur Transportwirtschaft und Kommunikation*, Hamburg.

Feldstein, M.S. (1964), "The Social Time Preference Discount Rate in Cost-Benefit-Analysis", *Economic Journal*, Vol. LXXIV.

Ferreira, L. (1997), "Rail track infrastructure ownership: Investment and operational issues", *Transportation,* 24, pp. 183-200.

Häusler, U. (1995), "Bildung und Funktion von Trassenpreisen für die Schieneninfrastruktur", *Zeitschrift für Verkehrswissenschaft*, 66, pp. 77-86.

Hedderich, A. (1996), "Vertikale Desintegration im Schienenverkehr", Band 11 der *Gießener Studien zur Transportwirtschaft und Kommunikation*, Hamburg.

Hylén, B. (1994), Access to Rail infrastructure -- Regulation and pricing. A State-of-the-Art Report from Ten European Countries. Summary of a Report to the Swedish National Rail Administration, Linköping.

Ilgmann, G. and M. Miethner (1992), "Netzstandardisierung und Preisbildung für die Fahrwegnutzung der künftigen Bahn", *Zeitschrift für Verkehrswissenschaft*, 63, pp. 203-229.

Nash, C.A. and J. Preston (1993), “Privatisation of Railways: The policy debate in Great Britain”, *Round Table 90*, *Privatisation of Railways,* ECMT, Paris, pp. 85-119.

Regierungskommission Bundesbahn (1991), *Endbericht der Kommission*, Bonn.

FRANCE

Luc BAUMSTARK
Chargé de Mission
Commissariat Général du Plan
Paris

Alain BONNAFOUS
Professeur
Laboratoire d'Économie des Transports
Lyon
France

Economic rail infrastructure tolls

SUMMARY

Paris/Lyons, January 1998

1. INTRODUCTION

The rail sector very quickly came to be regarded by economists as a typical example of a “natural monopoly”. In fact, like other networks, though often even more emphatically, it displays all the characteristics which tend to compromise the theoretical pact between the mechanisms of the competitive market and the optimum allocation of resources (in the Pareto sense). On the one hand, on such a market the break-even prices lack the qualities necessary to induce the economic agents to take optimal decisions. On the other hand, it is not possible to find a satisfactory price system that enables an optimum to be decentralised on the basis of the decisions of individuals.

The sector is, in fact, characterised by increasing returns resulting from multiple indivisibilities: functional indivisibilities such as the co-ordination of activities between upstream and downstream of production, for example, between infrastructure choices and commercial policy, but also technical indivisibilities such as that determined by the continuity of the network. To this there should be added various effects such as “economies of scope” which enable the operator, by diversifying the services offered, to reduce unit costs or important network externalities, both positive and negative, or, finally, investment irreversibilities. Thus, the rail sector accumulates the conditions for the appearance of what are commonly called *market failures*. This largely explains the sector's mode of operation in the early 80s and the major role played by government throughout its history.

However, the efficiency of this monopolistic organisation and the frequent public intervention in the sector have gradually been called into question on two counts. Firstly, it has been noted that the theoretical justifications for such an organisation apply only to part of the monopoly (mainly the infrastructure). Secondly, the very real advantages afforded by such an integrated structure have been offset by its no less real disadvantages (mainly its inability to exert sufficient pressure on production costs). Thus, the policies applied in this sector have sought to eliminate these weaknesses by reintroducing, wherever possible, various kinds of competitive pressure. Infrastructure access charges have become a keystone of the reforms. This question forms the subject of the first part of our paper.

There have been many modern theoretical developments in this field and our presentation does not seek to be exhaustive in this respect. However, there are two theoretical principles that deserve closer examination and these will be the focus of our attention in the second part[1].

The first principle states that infrastructure pricing cannot be separated from investment choices. The setting of "optimal" user fees presupposes the prior definition of a certain quality of service. For example, from the quality of service required there follows an acceptable saturation level which, in its turn, determines a co-ordinated policy of investment and demand management through infrastructure user fees. Hence, the theoretical economic toll is a composite toll, comprising a cost toll relating to the costs which can be directly allocated to the users of the network and an adjustment, sometimes called a pure toll, which enables the manager to produce the necessary quality of service by adjusting, as precisely as possible, the available capacity to the expressed demand (note that this second component only concerns the part of the network with no surplus capacity).

The second principle involves relating pricing to the difficult question of covering the fixed and joint costs which within the rail system, as in many other networks, represent a large proportion of total costs. The traditional theoretical solution based on marginal cost pricing, considered optimal, was to have these costs covered by the State budget. However, the perverse effects of systematic subsidisation, nowadays decried as public failures, mean that such a solution can no longer reasonably be regarded as optimal. Nevertheless, imposing budgetary constraint on the enterprise responsible for the infrastructure is not sufficient in itself to ensure the efficiency of the system. The pricing method and the formulas for allocating the fixed costs which, by definition, cannot rationally be allocated to any particular user, have definite implications for the social surplus which the system can produce. One of the main difficulties encountered in connection with pricing is that associated with the manager's technical and economic (but also political) ability to charge differential, even discriminatory fees.

In the last part, we compare these abstract principles with actual practice.

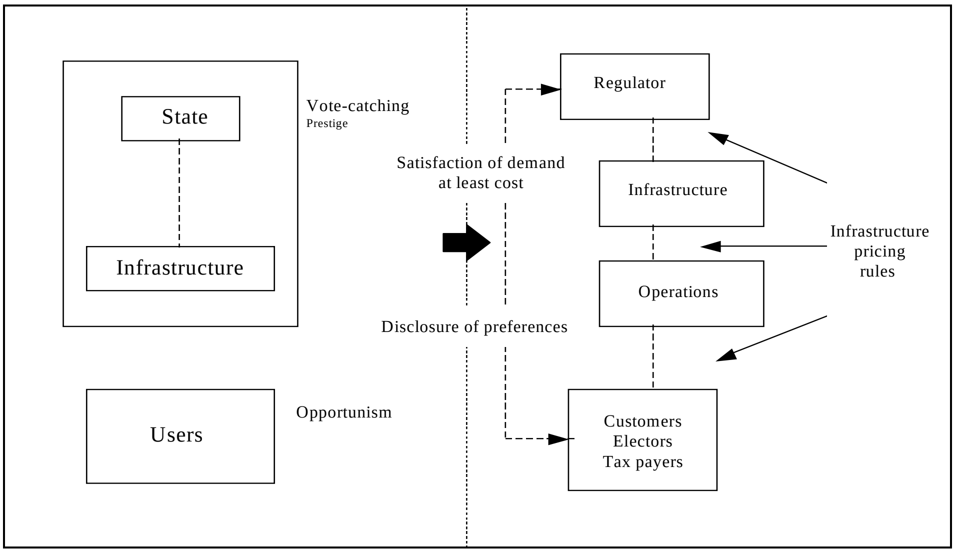
2. INFRASTRUCTURE PRICING, KEYSTONE OF INSTITUTIONAL INNOVATION

2.1 Network opening, redistribution of roles and pricing system

The idea of opening up the network is not a new one. Thus, networks have been identified in many different industries: the high-tension lines in the power distribution sector, optical fibre systems in telecommunications and delivery and collection in the postal sector, to which could be added numerous services such as the transport of electricity and information and the routing of mail. Although, as far as the network is concerned, it seems preferable to retain a monopolistic structure because of the economic characteristics of the infrastructure, this is no longer true of the services for which, on the contrary, competition seems possible.

In order to solve the problem of opening up the network, it is necessary to visualise how certain decisions which were taken internally within a single institution can be made the subject of commercial negotiations. The analysis must therefore focus on the types of relations that may exist between the decisionmakers, the private or public operators and the users. The configuration of the system and the actors is illustrated by the diagram reproduced in Figure 1.

Figure 1. **A traditional organisation chart**



It is thought that, under certain conditions, the intrusion of economic mechanisms into the integrated model will introduce transparent processes, thus obliging the various actors to disclose their preferences. This then leads to the explosion of an extreme situation in which the State produces the infrastructure regardless (or almost regardless) of the demand, of which it has only a faint idea. In this situation, the user benefits from a good without knowing its cost.

This can be improved upon by identifying intermediate levels. It is possible to distinguish between the production of the infrastructure and its management. Attention may also be focused on the behaviour of the agents and its determinants: the efficiency of the system as a whole depends on the efficiency of the means deployed to encourage and promote partnership between the economic actors and the authorities at various administrative levels. In short, it is a question of organising effective modes of interaction between the public and the commercial spheres.

On the one hand, it is necessary to establish relations between the State or, more generally, the authorities responsible for organising public transport and those directly concerned with the infrastructure so as to minimise infrastructure costs. This is part of the new theory of regulation which seeks to define more precisely the contractual framework within which these relations are to fit.

On the other hand, it is necessary to make sure that all the transport operators have access to the infrastructure on fair terms. This poses a specific problem with respect to the relations to be established between the infrastructure operator, public or private and the downstream service providers. Should or should not the monopoly be authorised to supply the market downstream? Can the sole and final responsibility for the allocation of timetable slots be entrusted to a body which itself also operates the infrastructure?

The determination of the prices at which the infrastructure manager opens up his network is one of the major considerations for the sector as a whole. These access charges are important since they determine the terms of competition between modes of transport. They must also ensure the overall efficiency of the system, that is, be sufficiently high for the upstream monopoly to be able to offer a suitably maintained infrastructure where the rail system is pertinent, but not so high as to bar the arrival of new entrants. More generally, pricing policy makes it possible, on the one hand, to encourage users to programme their services better (in the sense of making the best possible use of capacities) and, on the other, to steer investment towards ensuring that the system is developed and, in particular, the network expanded, so as to give the best possible cost-benefit ratio.

When he wishes to formalise the importance of the pricing system in the resource allocation process, the economist generally analyses the consequences of price distortion. Whatever the reasons, legitimate or not, for the distortion, the consequent loss of global surplus needs to be estimated. Within this theoretical framework, assuming an initial situation in which the pricing system is optimal, any change in prices not justified by a change in the cost of the production factors or by a sudden imbalance between supply and demand would inevitably lead to the misallocation of resources, a poor investment policy and, ultimately, less user satisfaction. In this case, the surplus available in the system would no longer be maximised.

This theoretical approach, even though it clearly relies on mechanisms less complex than they are in reality, provides valuable benchmarks for the orientation of supply and ensures cohesion between different strategies capable of serving the same purpose. The pricing system may then be understood as an information and co-ordination system, and if it offers a certain number of minimum qualities, the regulation of the system, being based to a greater extent on the rationality of the economic agents, will gain in efficiency.

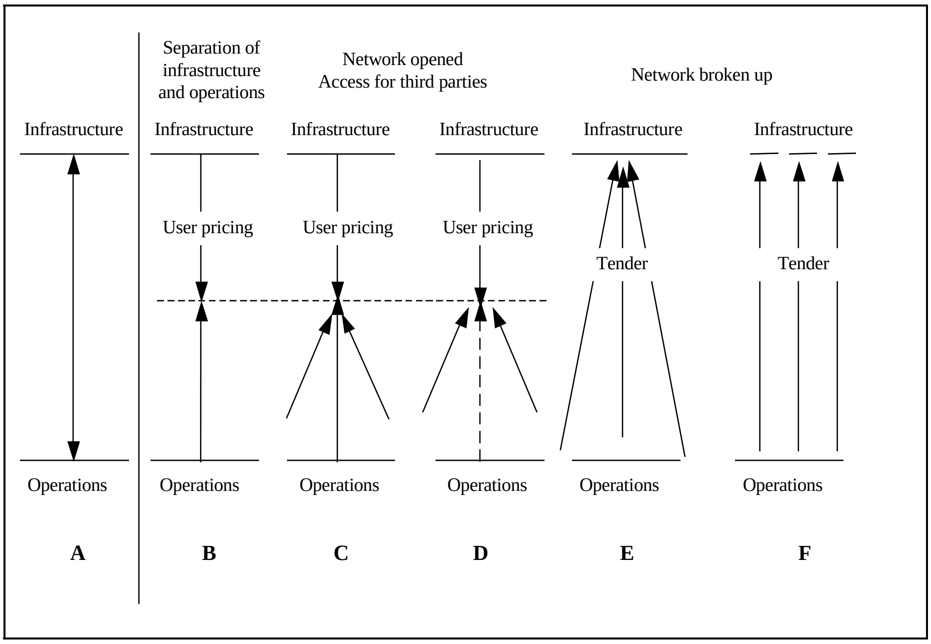
2.2. Degree of vertical separation in the rail sector and the question of access charges and user fees

These theoretical analyses underlie many of the recent trends in the European rail sector. They form the context for the adoption of European Directive 91/440, which first provided the impetus for the separation of infrastructure from operations. The Council of Ministers took a further step by issuing two new Directives 95/18 and 95/19. Before the network can be opened up, the conditions of network access must be established: licensing system, procedures for the allocation of existing capacities, coverage of infrastructure costs, network development, appeal procedures, etc. However, this Community drive is based on the reforms which many countries have already started to introduce. Although these reforms may employ different means, they apply the same basic principles, namely, the principles of decentralisation and the organisation of competitive pressure.

In practice, these widely introduced vertical separation policies are very diverse. This diversity contrasts sharply with a very homogeneous theoretical discussion of the question2. The main difference between all these systems appears to relate to the role of the State. Sometimes, though this is not the most frequent case, the State more or less organises the service, whereas the infrastructure is private3; in another configuration, the service may form the subject of a private monopoly, whereas the infrastructure remains in public hands (as in New Zealand). Under a third scenario, private companies operate a network in "partnership" with another private partner who owns the infrastructure (as in the United Kingdom).

The infrastructure user fee structure and its possible impact on the markets cannot be understood without taking into account the institutional configuration governing the management and operation of the network. A summary typology of the possible configurations is proposed in Figure 2.

Figure 2. **Typology of institutional configurations**



Situation (A) is the one that has been most frequently inherited from the past, namely, an integrated network occupying a monopoly position. As we have seen, adaptation, which may go as far as radical reform, involves creating levels of responsibility that give rise to commercial partnerships and to the introduction, with varying degrees of firmness of resolve, of the machinery of competition. Depending on the extent to which the network is broken down, several options are possible.

Assuming that the integrated monopoly is maintained, it is possible to imagine the introduction of competitive mechanisms similar to those in effect for certain urban public transport systems: invitations to tender for the management of the entire system lead to the appointment of a manager for a specified period. The integrated monopoly may be completely preserved (E). However, the network may also be broken up and tenders invited for each segment (F). The network may be divided up on a geographical or a sectoral basis, for example, with a network specialising in freight. In these various cases, the question of access charges does not really arise since the integrated monopoly is preserved: the bargaining takes place within the enterprise itself without involving any market services. The authorities may face other regulatory problems, but they will no longer have to deal directly with the question of access charges.

This does not apply to configurations (B), (C) and (D): although the integrity of the system as a whole is maintained (the infrastructure manager has authority over the entire network), the relation between the infrastructure and the operator or operators becomes distended. There are still many possibilities for reform, depending on the degree of separation between the infrastructure manager and the operator. Here the competitive pressure is exerted (with varying degrees of firmness) at the level of the infrastructure user by the (more or less) explicit means of diversification and dynamisation of the supply of services. However, in all three cases, it is a question of sending the operators a price signal calculated to influence their demand for the use of the infrastructure. The question of user fees will then be the determining factor.

Case (B) illustrates the lowest level of separation: third parties are denied access. Here, separation is intended simply as a means of clarifying the enterprise's accounts by dividing the expenditure items more clearly between those that relate to investment and maintenance costs and those that relate to operations. This model corresponds to the situation of those countries which, following the Community decision, preferred not to call into question either the integrity or the unity of their monopoly. Thus, the whole of the network continues to be run by a single operator. However, this system seems rather unstable and can be better understood as a staging post on the road to a more radical mode of organisation in which, separation having been made effective, third-party access can be encouraged. In the first case, pricing can remain a sort of accounting device by means of which the carrier is supposed to remunerate whoever undertakes to maintain and develop the infrastructure, that is to say, himself. Thus, the pricing rules are not really market mechanisms within which the parties endeavour to optimise their results. In cases (C) and (D), which are of more particular interest, the situation is quite different.

In case (C), the opening up to third parties is explicit but there remains a principal operator. Thus, the opening is still marginal and may be restricted to certain market segments, for example, the international market. This transitional situation may evolve into the much more competitive situation (D) in which, third-party access having been generalised, the historical operator is one actor among others. These two models lie at the heart of the two main scenarios for the implementation of the reforms undertaken in Europe.

We note, however, that the choice of system (C) does not protect the network concerned from sliding towards system (D). There will be de facto involvement of the international services and their demand for track will have to be reconciled with that of the national operator. Here, Community policy has clearly taken the same path as led road freight transport from progressive liberalisation of international transport to complete freedom of cabotage.

Consequently, the opening up of the network, however partial, raises the problem of how to regulate between the operator already in place and other potential operators. Even though the slot allocation rules may play a very important part, in the long run the calculation of access charges should prove decisive. Suppose, for example, that the allocation rules favoured the operator in place, say, through a practice involving “grandfather clauses” comparable to that observed in air transport, but that, at the same time, there was a heavy demand for slots, for international traffic, for example. The infrastructure monopoly could then take advantage of this excess of demand over available capacity by increasing the user price where the supply was scarce. If the use of the “grandfather clause” were too expensive for the historical operator, he could waive it and review his operating schedules in order to fall back on slots in better supply and the available capacity would then be sufficient. If it were still profitable to exercise the right despite the higher price, this would bring the infrastructure monopoly a return sufficient to finance the necessary investment in capacity.

In practice, this configuration might be relatively rare, but a similar problem might arise, including in the case of configuration (B), if the historical operator were to make, within the same space-time frame, requests for slots on behalf of different services (national and regional passenger services, freight services, etc.). If these requests were sufficiently well justified because of corresponding traffic demand, that would reduce to the expression of a need for investment in capacity on the sections in question. The historical operator might then have an interest in sending a “price signal” reflecting the shortage of available slots to each of the requesting services. It would be up to the latter either to buy because they were able to cover the cost or to transfer the demand to less busy routes or periods. This comes back to the above-mentioned choice between managing the demand and being able to finance additional capacity.

This alternative reduces to an infrastructure pricing principle which is well known but worth recalling in order properly to explore its conditions of application to rail transport.

3. INFRASTRUCTURE PRICING AND MAXIMUM EXTRACTION OF AVAILABLE SURPLUS

Infrastructure pricing may have more than one objective. The following presentation is founded on the idea that the basic principles should be established as a function of the collective efficiency of the system. Indeed, it would seem difficult to defend a pricing system which regularly deviated from that objective. Thus, this theoretical look at the problem suggests that there should be no segmenting of pricing between different, possibly contradictory objectives, namely, optimum use of the infrastructure, on the one hand, and the financing of its renewal or the pursuit of social or environmental objectives on the other. In fact, all these objectives can be reduced to the general objective of maximising the global surplus generated by the system.

On what principles should pricing be based in order to extract the maximum surplus from the system for sharing out among the various actors? This surplus can be identified only by relating it to the costs of the system and, in particular, the investment costs.

3.1. The concept of marginal cost and the optimisation of investment

Maximum efficiency obtained by marginal cost pricing in sectors with increasing returns is a decisive conclusion contributed by the theory of welfare economics. However, this theoretical result presents a number of problems. Apart from the undoubted difficulties of application, the concept is also open to theoretical objections.

In many cases, however, in discussing this issue, confusion is created insofar as the concept of marginal cost that is criticised is very often reduced to a user cost4. At first glance, this is hardly surprising. Understandably, since the marginal cost enters into the equation as the derivative of a total cost function, any cost factor that does not vary with production will vanish in the mathematical operation.

This classical presentation, without actually being false, can lead to inaccuracies with dangerous theoretical and practical consequences, simply because in this approach it is assumed that the investment is given and realised. In a manner of speaking, the act of investment is ignored as an optimising tool, on the same basis as pricing. However, the theoretical framework of the optimal allocation of resources cannot be satisfied with this reduction or with the solution consisting in transposing this short-term static model to the long term and treating the fixed costs as variables, thus causing them to vanish.

Indeed, taking the fixed costs into account presupposes an analytical approach quite different from that which consists in making them vanish. The fixed costs form part of the irreducible and fundamental linkage between the short and the long term. Leaving out that linkage amounts to evading the embarrassing questions raised by coverage of the costs thus incurred over several periods. From the standpoint of cost minimisation, it is well known that it is always possible to distinguish between directly avoidable costs5, costs that can be gradually absorbed and, finally, totally unavoidable costs. This distinction then makes it necessary to bring other concepts into play.

In particular, it is necessary to introduce a terminological distinction in order to avoid confusion. The variable marginal cost is a partial cost clearly distinguishable from the marginal cost, which has a different theoretical content: the partial cost relates specifically to a short-term situation in which only the variable costs are considered, while the concept of marginal cost relates to a situation in which it is assumed that the authorities meet the demand, if need be by carrying out a project that creates a discounted surplus greater than the total of the costs necessary to produce it. Thus, as many authors have stressed6, this concept is fundamentally linked with the investment decision.

Taking the long term into account requires the use of a programme situated upstream of the decision to invest. The fixed costs are not, strictly speaking, variabilised; however, the transport infrastructure is regarded not as a natural resource made available but as a good, to produce which the community must sacrifice resources.

Suppose that there are several technologies capable of meeting a given level of demand. Each of these can be characterised by a fixed cost, independent of the level of utilisation of the infrastructure and by a variable cost. The producer or the authorities have a certain number of possible solutions available to meet the demand. To each investment there corresponds a particular total cost function which depends on two variables, the investment I 1 and the level of production q:

$$C\left(I,q\right) = a\left(I\right) + b\left(I,q\right) \quad 2$$

Thus it is possible to distinguish between two marginal costs:

- the short-term marginal cost when q varies for a given investment level:

$$\gamma = \frac{\partial C\left(I,q\right)}{\partial q} \quad 3 \quad ;$$

- the long-term marginal cost when the investment varies: $\Gamma = \frac{\partial C\left(I,q\right)}{\partial I}$. 4

It is reasonable to assume that the producer (or the authorities) wishes to produce the good or service in question at the lowest possible price, if only to maximise the collective surplus. Thus, for each anticipated level of production, he will adopt the technology calculated to supply that good or service at the lowest price. Therefore, to each level of production there corresponds a minimum cost which itself relates to a particular total cost function belonging to the family of $C(I_i)$ 5. If it is assumed that I can vary continuously, the totality of these optimal points, which minimise the cost of production, will form a so-called long-term cost curve. We note that the curve obtained as a result of such a theoretical projection will be completely virtual. It defines a boundary of maximum efficiency of the production system at a given instant of time. Rather than an objective to be achieved, it is primarily a theoretical benchmark. It is a short-term curve envelope. This curve defines a functional relation between the long-term total cost and the quantity produced, on condition that each level of production is obtained with plant of the optimum size.

If it is assumed that the producer is able, at any time, to adapt his investment to the level of demand, there will be an infinity of short-term total cost functions. The system may be called upon to display its adaptability at any time. In this case, there is no contradiction between the short-term and long-term marginal costs. The short-term marginal costs are always defined, while the size of the plant is optimal for the production level in question.

To discard the hypothesis of permanent adaptability (divisibility) is to accept that the demand can be met by a technology or investment which is not necessarily optimal. In this situation, which is generally that of transport systems, there is almost inevitably a difference between the short-term and long-term marginal costs. There will then be several possibilities.

In the first case, that of underinvestment, the technology employed is not optimal since the short-term marginal cost is higher than the long-term marginal cost. In the second case, that of overinvestment, the short-term marginal cost is lower than the long-term marginal cost; the demand is met by an investment which costs too much for the use which is made of it. When, exceptionally, the system is optimised, the supply is precisely adapted to the demand, the two marginal costs are equal and, in this case, the marginal cost involves, over and above the short-term marginal expenditure, the optimal variations in the cost of the plant necessary to meet the demand.

> "*Under these conditions, and for practical applications of the theory of social returns, it is necessary to define the marginal cost of a specific service as the additional costs of all kinds (labour, energy, raw materials, depreciation, interest charges) resulting from the provision of an additional unit of that service when the existing fixed plant is precisely adapted to the volume of production in question*[7]."

Accordingly, the marginal cost to which pricing theory relates is defined at economic equilibrium and, consequently, with total utilisation of the fixed plant. This leads to a radical distinction between the marginal cost as described above and the partial cost. Thus, Marcel Boiteux considers it necessary, despite everything that may have been written on this subject during the last thirty years, to denounce, once again, the frequent identifying in common parlance of the marginal with the variable cost. He considers this "faulty" identification to be still "a major source of misunderstanding"[8].

This also means that the determination of the marginal costs has real economic significance only on the assumption that the infrastructure is optimally managed. For the concept to be pertinent, the maintenance operations and the investment renewal operations must be carried out at the optimal time and their cost must be minimised.

If the marginal costs thus defined are used for pricing purposes, the system will be coherent and optimal. This means that the consumers' choices will be optimally oriented since they will be encouraged to choose, among several ways of satisfying their needs within a given framework of constraints, that which is least expensive for the community. However, at the same time, the community must ensure that the demand thus expressed is always satisfied at the least possible cost.

This clarification is a necessary preliminary to the application of practical pricing tools. Within this context it is now clear that marginal cost pricing is future-oriented pricing. Accordingly, it should concern itself not with the previous cost of the infrastructure but with the use that the users will make of it. A knowledge of the users' preferences then becomes one of the keys to the global optimisation of the system from the standpoint of both economic calculation and pricing.

However, the balance between supply and demand thus achieved may give rise to very different situations on the rail system depending on the level at which the service quality is to be pitched. We have previously used expressions such as "meeting the demand" or "plant of optimum size". For example, should there be perfectly regular timetables with zero delays or would a reasonable amount of delay be acceptable if it permitted the more intensive use of capacity? Clearly, these questions prompt us to reflect on the quality of service to be provided which, in its turn, is also subject to a trade-off between costs and benefits.

3.2 The supply-demand equation or how to define an optimal quality of service

To deal with this problem, public economics traditionally relies on the notion of a mixed collective good, i.e. a good such that the quantity consumed can be distributed among the individual consumers (the good is therefore divisible): the consumption of a good by a user cannot be consumed by another, whereas certain other so-called quality characteristics remain indivisible because they concern all users. These mixed collective goods are subject to so-called congestion effects9. In the case of rail infrastructure, the indivisibility of its use can never be ensured inasmuch as two trains cannot share the same "slot" at the same time. The slots are shared between the different users. The use of infrastructure, from this point of view, is divisible. On the other hand, the quality of service, which depends on the reliability of the train timetables on a congested part of the network, is the same for all the users at a given instant10. This collective good is called mixed because it has two fundamentally different characteristics, the first -- access to infrastructure -- being divisible, the other not. Congestion is a particular case of an external effect *where the reasons why people cause and suffer it are linked with the consumption of the same service.*

In order to optimise the system and achieve the maximum available surplus on the infrastructure, the authorities must both determine the optimal level of investment and, through good pricing, manage the level of demand. Investment and level of demand are the two factors of the quality of service offered which different users value in varying degrees.

In order to isolate the problem posed by this linkage, let us consider an infrastructure, the cost of using which is very low and may therefore be deemed negligible11.

Let us suppose that all the users i liable to use the infrastructure place different values on the use of this good, for a given quality of service. We then obtain an inverse demand curve which we will denote by $P_i(Q_i, g)$, where Q_i represents the utilisation of the infrastructure by the individual i, and g an indicator of service quality.

The quality of service offered on this infrastructure depends on two parameters: the number of users on the infrastructure and its physical characteristics, which depend on the level of investment k. This level of investment varies with time as a function of the outlay made by the operator, $\psi(k)$ 6.

Taking into consideration several time periods t, we write $P_{it}(Q_{it},g_t)$, with $g_t = (Q_{it},...,Q_{nt},k)$ the quality of service at period t. We then suppose that each user integrates this quality of service, $P_{it}(Q_{it},g_t)$ being the value that the user i

accords to the use of the infrastructure at a given level of quality g. This value may evolve over time.

The individual surplus is given by the difference that exists between the value that the individual attributes to the use of the infrastructure and the cost which that use represents for him, remembering that in this case the user cost, in its strictest sense, is seen as negligible. This cost is therefore limited to the deterioration in quality of service suffered by the user, due to the utilisation of the infrastructure by others. This cost is all the more important the greater the number of users and the weaker the investment. If we take D_t as the deterioration of service quality at time t and V_{it} the value that the user i gives to this deterioration, the surplus of user i for each period t is expressed by:

$$P_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right)$$

Now let us assume that the operator is seeking to maximise the social surplus which this infrastructure is capable of generating. For each individual i and for the entire length of time in question, we have:

$$\sum_{t=1}^{T} \left(\int_0^{Q_{it}} P_{it} . dQ_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) \right)$$

The global surplus is obtained by summing over all the users and subtracting the investment cost. This calculation leads to the following expression:

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \left(\int_0^{Q_i} P_{it} . dQ_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) \right) - \psi(k)$$

The maximisation of the surplus depends on two variables, the utilisation of the infrastructure and the level of investment.

The first n partial derivatives have the form:

$$\frac{\partial \bar{S}}{\partial Q_{it}} = P_{it} - V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) - \sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\partial D_t}{\partial Q_{it}}$$

The relation for the investment is written as follows:

$$\frac{\partial S}{\partial k} = -\sum_{t=1}^{T}\sum_{i=1}^{n} V_{it} Q_{it} \frac{\partial D_t}{\partial k} - \Psi'(k)$$

7

Thus, the first-order conditions lead to the following two relations:

$$P_{it} = V_{it} D_t\left(Q_1 \ldots Q_i \ldots Q_n, k\right) + \sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\partial D_t}{\partial Q_{ti}} \quad \textbf{(1)} \text{ and}$$

$$\Psi'(k) = -\sum_{t=1}^{T}\sum_{i=1}^{n} V_{it} Q_{it} \frac{\partial D_t}{\partial k} \quad 8 \qquad \textbf{(2)}$$

Equation (1) establishes the principle of the pricing rule. The first term of this expression represents the "cost" accepted by the user i resulting from the loss of quality of service associated with congestion. The second term represents the value of the quality of service lost by all the users because of the last utilisation by the user i: this last term represents the external cost of congestion for which the latter is responsible.

Using classical pricing terminology, the optimal pricing is such that the price corresponds to the sum of the private marginal cost and the social marginal cost. The first term is already borne by the user. The toll to be applied should therefore correspond to:

$$\sum_{j=1 et j \neq i}^{n} Q_{jt} V_{jt} \frac{\partial D_t}{\partial Q_{ti}} \quad 9$$

If we now consider the second equation (2), which incorporates the investment dimension, it indicates that the capacity should be developed until the marginal investment is equivalent to all the congestion costs avoided.

Finally, and this is crucial, the optimum in terms of social efficiency will only be ensured if these two conditions are satisfied. The golden rule of marginal cost pricing consists of these two conditions, which theory suggests should be kept together. In common parlance, it could be said that this dual logic consists in establishing a price which, wherever there is saturation of the network, ensures either that demand is sufficiently well managed for the available capacity to be still sufficient or that the cost of the necessary investment in capacity can be covered. However, the quality of service production function may have different characteristics depending on the importance accorded to the intensity of the demand. The optimisation procedures will then be more or less determined by scarcity phenomena.

The congestion toll described above may take several forms, depending on the degree of divisibility of infrastructure use. The less the divisibility, as in the case of road transport, the more the intensity of use will affect the quality of service. But the greater the divisibility the less perceptible this effect and the more pricing should be oriented towards scarcity management systems, as in the case of car parks. The analysis should therefore be focused on the technical and economic relationship between the infrastructure and the uses of that infrastructure.

The rail system does not readily lend itself to this type of analysis. The existence of timetable slots seems to give credence to the idea that the use of the infrastructure is totally divisible and that there is therefore a strong user rivalry among the various consumers. Thus, it is theoretically possible to establish a market in user rights, for example, by auctioning slots, and thus reconcile supply and demand. It seems that this might be practicable on railway lines dedicated to similar kinds of traffic since, in this case, the slots auctioned would tend to be homogeneous. However, the only configuration that would seem to lend itself to this exercise is that of lines specialising in freight and open to several carriers, which could only apply to the lines, still to be organised, of a trans-European rail freight network. The first "freight corridors" to be established are not organised along these competitive lines.

In most cases, competition remains latent and overshadowed by slot allocation rules which predate the reforms. However, it is very real wherever network saturation is a problem, which brings us to the difficult question of capacity.

Rail capacity on part of the network can be very roughly and provisionally defined as the maximum possible number of movements that can be handled per hour. The reality, however, is much more complex.

On the one hand, capacity is determined by the characteristics of the infrastructure itself: the configuration of the lines and sets of tracks, the existence of community lines, switches, track intersections, the speed permitted by the design of the track, gradients, etc. Moreover, the capacity also depends on the utilisation of the infrastructure: type of trains (length, axle load, etc.), speed, number of stops, etc. The train schedule is then decisive and its organisation will have an impact on the network's effective capacity, which will be all the more considerable the more heterogeneous the slots concerned. The capacity will then depend on the train schedule as it has been drawn up. The scheduler and the rules of arbitration on which he relies will thus play an important role, a role which, in railway tradition, has not been much influenced by the considerations of the economist who, for his part, is keen to maximise the surplus that can be extracted from the infrastructure. To that end, he analyses the economic advantages which the consumers derive from using the infrastructure. For example, a decrease in the time taken by a train to travel between two stations may lead to a reduction in the technical capacity of the infrastructure by eliminating slots, but that reduction in capacity may be justified because it results in a net gain in surplus production. Conversely, the economist lacks mastery of the complex relations between the characteristics of the infrastructure, the possible operating programmes and the response of the final demand.

This difficulty involves both infrastructure pricing and project evaluation. Whether it is a question of calculating a developing marginal cost or the return on an increase in capacity, it should be kept clearly in mind that such an increase can very often be achieved in different ways: an additional junction, flyovers for avoiding track intersections, higher speed limits, passing track en route or in the station, improved block arrangement, better power supply, safety and traffic control installations, restoration of alternative routes, etc. As with the roads, the optimum infrastructure capacity is not a technical but an economic factor which, however, for rail forms part of a complex universe of technical solutions.

Here, we have one of the explanations for that common tendency of railway reforms which "separate" infrastructure management from transport production. The unified monopoly operator or separate operators not being omniscient, whatever the theoretical optimisation models may suppose, have not efforts been made to improve the efficiency of the system by giving the carrier the opportunity to make the best possible use of the signals sent him by the infrastructure manager and vice versa? Thus, it is for the carrier to interpret the infrastructure user price signal and for the infrastructure manager to interpret the demand signals he receives from the carrier. These "interpretations" could be based on the teachings of public economics.

However, to the above-mentioned difficulties we should add another associated with a familiar feature characteristic of most rail systems, namely, the presence of very considerable fixed costs and modest marginal costs over most of the network, i.e. wherever there are no saturation effects. The principles we have noted then lead to pricing which can result in a considerable deficit which it is the community's responsibility to meet. Whence the now acknowledged need to take this "budgetary constraint" into account.

3.3. Allowing for the budgetary constraint

The doctrine according to which the producer price of industries with diminishing costs should depend only on the marginal operating costs, and the rule according to which the authorities should cover all the fixed costs from taxes, first made an explicit appearance in the railway literature of the late nineteenth century. The French tradition of the economist engineers of the École des Ponts et Chaussées made a big contribution to these developments and their application. The debate became more animated following the appearance of Hotelling's article[12] in 1938. He concluded that the deficit resulting from the application of this global pricing principle should be financed by global taxes which, like the taxes on rent charges or inherited income, are supposed, in theory, not to affect the marginal behaviour of the economic agents.

Nevertheless, there may be a conflict between this theoretical viewpoint and another scarcity phenomenon, the public finances. Moreover, it assumes that resources are optimally allocated by the public operators, which is not necessarily the case when they believe themselves to be free of all financial constraints. It is to correct these dysfunctions that an entire theoretical school has devoted itself to justifying the addition of a budgetary constraint to marginal cost pricing. There are several ways of understanding this approach.

For some, this concern for a balanced budget is, in the words of Serge-Christophe Kolm, no more than the obsession of "*a short-sighted and narrow-minded accountant, who knows nothing of economics, trying mistakenly to transpose the criteria of the private to the public sector*[13]." By refusing to consider the main analytical conclusions of welfare economics, the supporters of the balanced budget deny themselves the possibility of satisfying, with the tools of economics, the demands of the public interest. Without rebutting this position directly, Mark Blaug has made it the target of some equally critical remarks which reflect the difficulties created by this pricing principle. Thus, according to Blaug, what the Anglo-Saxon authors call the French school[14] has trouble in taking the existence of deficits into account: "*The characteristic feature of the French contributions to the literature (on marginal cost pricing) is a total inability to take into account the problem of deficits in the diminishing cost industries which, indeed, hardly receives a mention*[15]."

For others, there are fewer disadvantages (i.e. less loss of global surplus) in distorting optimal pricing to control the deficit than in leaving the latter to drift while seeking to adhere strictly to the optimal pricing principle. Thus, in France and Europe, there has been a slow swing of the pendulum: the arguments in favour of marginal cost pricing no longer convince the authorities, who are more concerned about the financial situation of the public corporations, especially that of the railway companies which are steadily losing modal share, as well as about the unfortunate effects which the systematic covering of the deficits is having on the management of those enterprises.

Driven by the structural difficulties of the public finances, this trend is also based on the failure of a pricing system which requires transfers between taxpayers and users to ensure that the cost of the service to the community is explicitly weighed against the interests of those who use it. The gap between those who benefit from the system and those who finance it leaves a space within which the economic agents can conceal their preferences. The deficit subsidies may also hide and hence permit inefficient operation. Now, if there were no strong incentive to seek the minimum average cost, the willingness of the State to close any gap between the marginal cost and the average cost would result in enormous waste, which would be all the more enormous as it would probably be invisible and almost undetectable[16]. Thus the balanced budget constraint is aimed at an efficiency deficit which goes far beyond the traditional criticisms levelled against the champions of marginal cost pricing concerning the difficulties of evaluating the marginal cost and the technical, political and institutional barriers to implementation.

The following list of criticisms, without being exhaustive, will serve as an illustration: the pronounced indivisibilities of the infrastructure would lead to "sawtooth" pricing incompatible with long-term decision-making by the economic agents; the lack of rules for allocating certain cost elements would make competition impossible because of the existence of cross-subsidies; differential pricing on the network would bring into question an entire spatial standardization system; the practice would make it necessary for both users and authorities to gather, at great expense, information on the competitive structure of the market, on externalities, on the elasticity of the demand, etc. and, in short, the additional costs which the public would have to bear to implement these pricing systems would be out of all proportion to the advantages which they are supposed to bring.

As Vickrey points out, we are constantly on the horns of a dilemma from which it is difficult to escape completely. On the one hand, theoretically, the application of marginal cost pricing ensures that infrastructure utilisation is optimised but, considering the financial scarcity constraint, we then deprive ourselves of information on the real value of a new project or, if the project has already been carried out, about whether it is still worth operating. On the other hand, with the balanced budget constraint, we can be sure that the project or its operation are worthwhile, but we do not know whether the infrastructure is being utilised to best advantage[17]. More generally, by employing crude regulatory mechanisms, by excluding a number of users and by eliminating certain activities, the application of such a rule might lead to a serious loss of social efficiency[18].

All the solutions proposed for introducing a budgetary constraint into the pricing system must face this radical criticism: the allocation of the fixed costs to a particular user or a particular use remains largely arbitrary. Thus, the calculations form the subject of endless discussions about the relevance of a particular distribution scheme. It might seem that subtle cost accounting could reduce the proportion of non-apportionable costs and thus eliminate the problem. However, these methods can always be debated and offer no solution for a usually still significant residuum.

Now, from the standpoint of the optimal allocation of resources, the pricing system should not be mainly concerned with distributing the costs but, more fundamentally, should favour the achievement of a surplus. The objective is much more ambitious. Leaving aside the utilisation costs, for which the allocation procedures do not, in theory, pose any particular problems, it is clearly the availability of the infrastructure that must be reflected in the scales and not its effective consumption. Then, the efficient mode of contribution is not to seek to allocate costs but to find a means of extracting the surplus needed to finance the infrastructure while ensuring that the surplus is achieved. From the moment that this surplus exists (it is the role of public economic analysis to locate it), there exists a pricing system to bring it to light.

It is the segmentation of the demand which, in this respect, makes a vital contribution to economically efficient pricing. The theory of surpluses leads to precedence being given to pricing systems based on this latter principle. In practice, it may lead to the adoption of the monopolists' rule to the effect that the best pricing consists in imposing the charges that the traffic will bear.

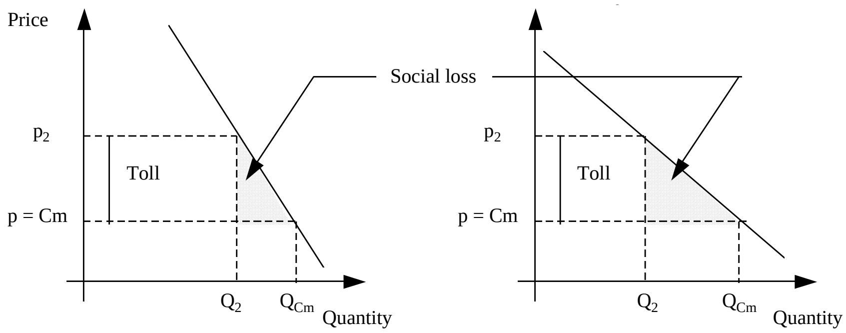
In fact, the question of the social loss involved in pricing that deviates from the marginal cost, like the question of the deficit and the limitation of its perverse effects, makes sense only if price unity is assumed. As soon as a discriminatory approach is adopted, the difficulty may disappear. Price discrimination, which takes into account the response to prices of each segment of demand, then makes it possible to increase the global surplus since the number of users will increase, while ensuring better coverage of the costs for the producer. The introduction of socially efficient price discrimination turns pricing completely upside down. It should be distinguished from the pricing principle, often encountered in the literature, that everyone pays his share, which makes even less sense inasmuch as the allocation of some costs is arbitrary. It forms part of another approach which authorises any use of a good calculated to generate a positive net surplus. Under this condition, not only will any discriminatory pricing system be neutral from the standpoint of the optimal allocation of resources, but it will be totally justified from the standpoint of the community since it will enable a surplus to be generated. Thus, the first principle considered to characterise optimal pricing gives way to another principle.

A pricing system is deemed to be economically satisfactory if the operator procures for each user a share in the absolute utility of the service sufficient to constitute an effective incentive for him to use the infrastructure. The contribution to this service is then said to be fair as long as, for each consumer and for each use, it does not exceed the net value of the utility he derives from it which, it should be recalled, must be positive. Although often disputed[19], this approach, based on tapping into a surplus, has a definite advantage in relation to the problem of financing which, as Jules Dupuit suggested, consists in “demanding as the price of the service provided not what it costs the provider but an amount commensurable with the importance attached to it by the one for whom it is provided[20].

This basic principle remains very theoretical and there are major difficulties to be overcome before it can be applied. However, the tools do exist. In 1956, Marcel Boiteux[21] proposed a solution which marked an epoch in the history of economic thought. The literature has even associated the name of the author with that solution so that it is customary to speak of Ramsey[22]-Boiteux pricing. Anglo-Saxon authors also refer to this principle as the “Inverse Elasticity Rule”. Tradition has it that this seminal article offers a general solution of the problem of the production and the Pareto-optimal pricing of a public monopoly obliged to balance its budget within the context of a competitive economy.

Here, then, the decisive concept of elasticity of demand makes its appearance. Thus, if the enterprise is considered to have several markets at its disposal, observance of the budgetary rule which requires deviation from the marginal cost will lead to the prices for each market being so determined as to make the mark-up between the consumer price and the marginal cost proportional to the inverse of the price elasticity. In practice, this comes down to saddling the goods or services for which the demand is relatively inelastic with a larger share of the deficit. This pricing practice harks back intuitively to a mechanism well known in the field of taxation: the loss for the community resulting from the imposition of a tax (that is, in this case, the reduced consumption of a good consequent upon an increase in price) will be the greater the more elastic the demand.

Figure 3 **Social loss and elasticity of demand**



Assuming a monopolistic market on which the demand is very elastic, the slightest variation in price on this market will be reflected in substantial changes in the pattern of consumption. Thus, the so-called Boiteux rule consists in taking advantage of different relative market situations. The steeper the slope of the demand curve on a market, the lower the elasticity and the more limited the social loss resulting from a deviation from the marginal cost. Conversely, on a highly elastic market, the slope of the curve will be flatter. In this case, a deviation of the price from the marginal cost will be reflected in a heavy social loss.

Thus, when by necessity an enterprise has to cover the whole of its costs and hence, in the case in question, when it is obliged to deviate from the marginal cost, this theoretical demonstration makes it possible to justify placing the strain on the consumption which is the most inelastic. When the demand is relatively inelastic, the deviation of the Ramsey-Boiteux price from the marginal cost will be small and hence the deviation at the optimum will be minimised. The mark-ups will be greatest on the least sensitive demand. Thus, this method of pricing seeks not to distort the price signal sent to the most sensitive users in order that they may not significantly modify their pattern of consumption and to levy the charges on the less sensitive users who will not reduce their consumption more than slightly relative to the social optimum, even if the prices are raised.

Thus, returning to the allocation of non-apportionable fixed costs, where a monopoly can rely on several products it should parcel out its fixed costs according to the sensitivity of the demand. Prices are raised sharply where the demand is not sensitive and reduced where it is. The quantity consumed on each market remains as close as possible to the consumption which would have been observed in the first-best case. The optimum obtained maximises the social surplus subject to the constraint of a balanced company budget or, if this constraint seems too harsh or inaccessible, by assigning in the Ramsey-Boiteux optimisation programme a scarcity coefficient which overestimates the collective utility of the public contribution.

As Boiteux has himself been pointing out since 1956[23], to the difficulties of application of this pricing method there must be added the unrealistic assumptions of the model. At the same time, the author questions whether the practical application of the results obtained is of any real interest. In fact, the application of this rule poses a number of problems, in particular, by requiring a knowledge of demand elasticities. More generally, the necessary hypothesis of an omniscient economic supremo is obviously unrealistic. The fact that it is a question of pricing the use of the rail infrastructure and not the transport service itself is a further complication. The carrier is then the source of the demand, in the sense that he needs slots and he himself responds to a combined demand: that of the users or shippers and that of the transport organising authorities who „purchase" qualities of service from him. Thus, the effects of infrastructure pricing on the final demand are linked with the pricing principles which the carrier himself applies. For the system to be coherent and economically efficient, it is very likely that the two operators will have to price in accordance with similar principles, but this linkage has still to be subjected to a complete analysis.

This brings us to an essential conclusion concerning the pricing system which Maurice Allais considers to be one of the key elements distinguishing his theory of the markets economy from the standard model[24]. Within this more general theoretical framework, which seeks to maximise the collective surplus, the deficit constraint can be removed provided there is no objection to questioning price unity. That is an intuition already clearly expressed by Jules Dupuit[25] in his time. If the surplus generated by an infrastructure is greater than the cost of putting it in place, then there necessarily exists a pricing system capable of tapping into this surplus to obtain the sums needed to finance the project while maintaining maximum social utility. The achievement of the surplus becomes the challenge of the pricing procedures. It can justify the transition from a logic of differentiation based on costs alone to a logic of discrimination based on segmentation of the demand. It is not a question of differentiating prices solely on a user pays basis, which makes little sense since, once again, a portion of the costs cannot be distributed in accordance with this principle. The discrimination of the demand should be based on the principle of he pays who can, especially when that is the only way of obtaining a return on projects with a positive global net utility. Otherwise, if the project is not financed out of the general budget, a type of solution now considered best avoided, it will never materialise.

We then enter a pricing universe that is more complex but still regulated by marginal cost in which it is less a matter of minimising the effect of a deficit linked to optimal pricing as of seeking to maximise the surplus by differentiating the pricing, segmenting the customer base and, finally, obliging all the economic actors to disclose their preferences. These principles now need to be compared with the actual results of implementing the most significant recent reforms.

4. APPLICATION OF PRINCIPLES AND NETWORK CONSTRAINTS

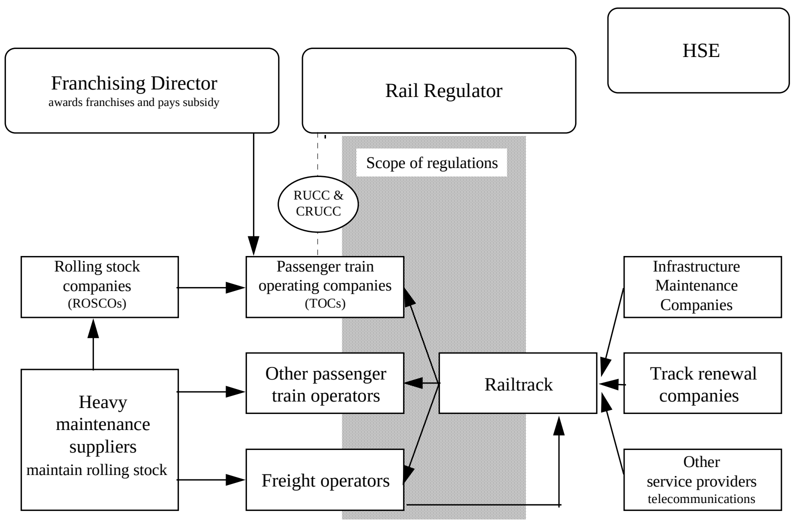
In this final part of our paper, we shall compare the few theoretical principles that can be derived from economic analysis with the practices of networks which, as a result of having been reformed, should have solved the infrastructure pricing problem. Accordingly, we shall examine, in turn, the British, German and French cases.

4.1. The British experience: making a surplus[26]

The restructuring of British Railways began in the early 80s. This long process, intended to improve the efficiency of the sector, led the authorities gradually to modify the enterprise's internal organisation. On 1 April 1994, the reforms took a further step forward with the entry into force of the Railways Act which prepared the ground for a phased privatisation of the sector by programming, in particular, for the separation of operations from infrastructure. The enterprise's transport services have thus been separated into smaller entities which, from the outset, the authorities have made clear they intend to privatise. As for the infrastructure, it has been assigned to a new, private-law company held by the State, Railtrack. This company, still a monopoly, retains operational control over the traffic, allocates capacity and, above all, is responsible for pricing use and determining the fees to be paid by the various operators to ensure that the costs are covered27. The company, which initially was left in public hands and even benefited from investment subsidies, was privatised in 1996. BR's passenger transport business has been split up into 25 separate entities which have been placed under private sector control by introducing a franchising system. The freight business has all been sold off to the private sector and opened up to competition. Thus, Railtrack buys services from and sells them to a range of operators.

At the same time as splitting up BR, the Railways Act established a powerful regulatory system based on three bodies. The first, the Office of the Rail Regulator (ORR), is mainly responsible for supervising infrastructure access and pricing. It establishes the rules of competition and oversees their application, especially in the interests of the customer. The task of the Office of Passenger Rail Franchising (OPRAF) is to grant franchises and supervise the correct application of the terms and conditions by each franchise holder, in particular as regards the consistency of the services actually provided. Finally, the Health and Safety Executive (HSE) ensures that the safety regulations are observed. It issues rules governing the design, construction and operation of rolling stock, infrastructure and equipment.

Figure 4. **The British reforms**



Source: ORR Report, 1996-97.

Each operator signs a track access agreement with Railtrack. Two sectors should be distinguished. Initially (7-15 years), the franchised passenger line operators have been granted the access rights necessary to provide the services stipulated in their specifications. On certain routes they enjoy protection which will eventually be withdrawn. The existing freight carriers have been granted initial rights to enable them to satisfy their present customers. Apart from the time slots allocated in connection with these rights, there are others which are open to competition.

The splitting up of the network immediately posed the problem of the allocation of costs among the various activities. It was decided that the costs should be allocated to the various sectors in such a way that each sector bore the costs of the fixed assets and personnel of which he was deemed to be the principal user. The basis for this pricing is that it must be sufficient to ensure Railtrack a certain return on its assets.

The rail operators pay Railtrack infrastructure user fees intended to cover the network utilisation and signalling costs and the cost of supplying power, where appropriate. Thus, the overall aim is to pursue a balanced-budget pricing policy which also takes into account the ability to pay of the applicants for time slots. Accordingly, it is not possible to speak of Ramsey-Boiteux pricing because of the special terms granted to the franchise operators.

In fact, the pricing system differs depending on whether one considers the operators who provide services under franchise (subsidised services) or the operators who purchase time slots on the network. The companies which operate passenger lines under licence are in a special situation since in this case the structure and the level of the access charges applicable are directly controlled by the regulatory body. The access charges applicable to the passenger lines operated under licence[28] are based on a cost allocation study that uses the concepts of avoidable[29] and additional[30] costs.

In the event of a time slot purchase, Railtrack is free to negotiate its prices, although the contract must be approved by the Regulator. The fees are negotiated but subject to approval based on the principles established by the Office of the Regulator.

The general principle requires that the fee structure should not deviate too far from the value of network access for the users and that it should enable Railtrack to recover all the costs actually incurred in connection with the transport of goods, to which there should be added a possible contribution to cover the joint costs shared with the passenger services. Thus, the pricing rule should be such that the minimum price is not less than the avoidable costs occasioned by the service concerned. The price should be less than or equal to the costs which Railtrack would incur if the operator were alone on that portion of the network and then had to assume all the costs. The price should not, when the various cost factors are taken into account, differ appreciably from one user to another.

In addition to the transparency of the charges being more difficult to achieve for freight than for passenger transport and it being difficult to determine precisely the share of the cost directly apportionable to a particular service, the allocation of the joint costs is a real headache.

The reforms have been the target of two main criticisms. Some consider that Railtrack, by setting low prices for freight transport, for which it is competing, may seek to shift the fixed costs onto the passengers and thus improperly obtain public funding to its advantage. Others, on the other hand, consider that the draconian safety regulations, though reasonable in the case of passenger transport, are less justified in the case of freight transport and, accordingly, that there has been a transfer from the freight to the passenger sector.

However, in 1995, before Railtrack was privatised, after examining the access charges applied to the franchised passenger services, the ORR concluded that they were too high and more than the operator really needed to fulfil his infrastructure renewal obligations. The Regulator imposed corresponding price reductions, thus transferring the productivity gains to the licensed operators. The charges are to be reviewed in the year 2000. Another object of criticism is the charge structure, the fixed portion being considered too high (91 per cent). This pricing is not conducive to the rational management of resources. In fact, this approach precludes the introduction into the access charge calculations of differentials in terms of rush hours or the economic value of the slots. The costs considered here do not incorporate such externalities as noise or air pollution.

When the pricing rules were being drawn up, there was a keen debate between the advocates of a commercial strategy and those who favoured a more managed approach. The arguments are important inasmuch as the same debate is now being conducted at the European level. After holding numerous consultations, the ORR concluded that it was better to place the method of calculating the infrastructure user fees on a commercial negotiation basis so as to give the operator a chance to attract all the economic agents capable of paying at least the directly apportionable costs. Thus, the main aim is to give the infrastructure manager the means to induce the maximum possible number of operators to use the network. Clearly, then, the objective is to achieve and engineer a surplus. This approach has been much criticised. The owner of the infrastructure can engage in cross-subsidisation and favour one party or another without necessarily ensuring the complete opening up of the network

4.2. The German experience: covering the costs

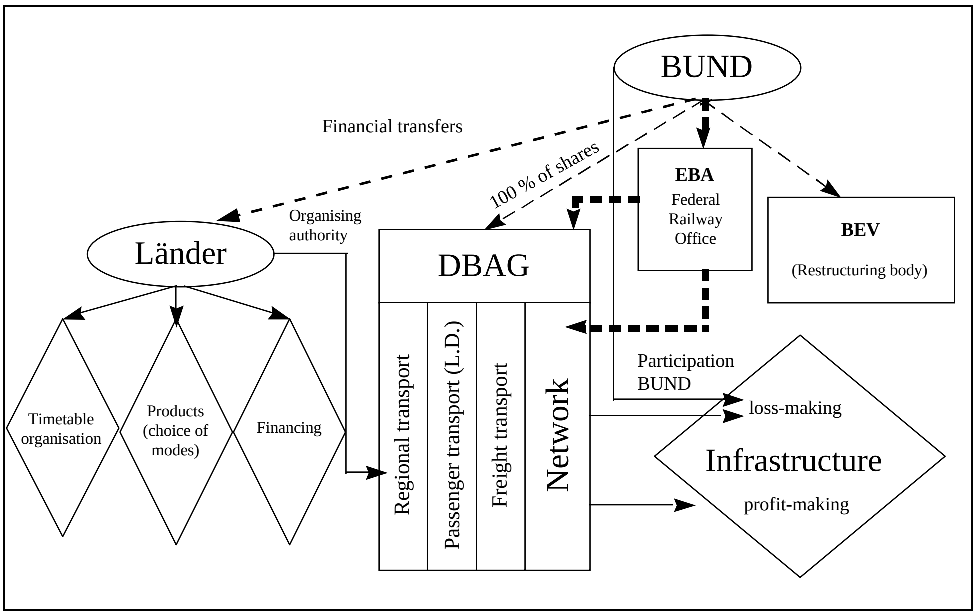
As in other countries, the amount of federal subsidies granted to the rail sector in order to balance its budget and guarantee its borrowings had become considerable while, at the same time, rail's share of total traffic was being continually eroded. Reunification only worsened the crisis and, in 1993, the enterprise (made up of the DB and the RB) had debts of more than DM 67 billion while the Government was anxious to support the development of the sector. The reform of German railways was speeded up, profoundly transforming the rail transport situation across the Rhine. It opened the way for ever keener competition on the railways, on the one hand in the short term, by placing the historical rail operator in the position of a service provider in its negotiations with each Land and, on the other, in the long term, by opening up the network to third parties. The reforms entered into effect, following an amendment to the Constitution, on 1 January 1994.

Thus, the authorities gradually moved towards a vast controlled and concerted structural reform of the railway company. The Central Government (*Bund*) set up a private joint-stock company, the DB AG, in which at present it holds all the shares but which is supposed to disappear in 1999. The rail system is organised in four independent sectors of activity: regional passenger transport, long-distance passenger transport, freight transport and infrastructure management. These sectors are eventually to be privatised. However, the Federal Government will remain the principal shareholder in the *Fahrweg* (infrastructure manager) in order to retain control over investment policy.

As in the United Kingdom, the reforms are based on new institutions. On the one hand, the Federal Office of Railways -- the *Eisenbahnbundesamt* (EBA) -- has been set up to ensure the necessary co-ordination and take care of the general missions of the railways. It authorises operations, certifies equipment and organises work on the infrastructure. It plans the work on the federal rail network, ensures that tendering procedures comply with the law, grants licences, applies investment financing agreements, commences prosecutions and settles disputes. On the other hand, the *Bundeseisenbahnvermögen* (BEV), another federal body, is responsible for clearing the debts of the former company and for administering staff and pension costs as well as financial and property charges. Central government has taken over the costs of the staff of the former DB with civil servant status, who can now be made available to the DB AG for employment under ordinary market conditions.

At the same time, to meet the increasing costs of regional rail transport, the central authorities have transferred to the *Länder*, with their agreement, the responsibility for organising and financing regional transport operations. Thus, the latter will henceforth find themselves in the position of organising authority. They will receive financial transfers from the Bund to enable them to perform this task31.

Figure 5. **The German reforms**



Infrastructure pricing is an important component of these reforms. The Act authorises access for third parties and thus transposes into domestic law the principles of Directive 91/440. The operating subsidiaries of the DB AG pay the infrastructure manager for the use of the infrastructure. These charges are published in a catalogue which tells the operators exactly what price they will have to pay for the whole of a journey, depending on their requirements. In July 1994, DB Infrastructure published an initial price schedule applicable to all users of the network. The system was chopped up into 4 000 sections with well defined characteristics on the following principles.

The prices are identical as between the DB AG and third parties and must be identical for requests with similar requirements. The differentiation to be found in the catalogues is based on objective criteria. The prices depend on the category of line, the damage potential of the equipment used, which largely depends on the type of use, the required regularity, the volume of purchases and the length of time for which the slot is used. The basic tables are compiled on the basis of four criteria, namely, the quality of the track (essentially the permissible speed), the traffic potential (principal characteristics of the rail links) according to the type of service requested, the wear and tear (based on a variable cost analysis) and the planning quality. This last item relates to the quality of service demanded, the reliability (punctuality) indicator determining the room for manoeuvre left to the infrastructure manager. This indicator is expressed in the form of a percentage representing the margin which the DB is allowed relative to the theoretical journey time. Using these criteria, it is possible to construct the reference table shown in Table 1.

The notion of quality introduces an element of economic demand management. Here, the aim of differentiation is to define a scale of operator's requirements. In choosing a quality of service, the operators disclose their preferences. However, there are strict limits to this mechanism since the percentage is fixed for each category. It would be different if the operator could choose a level of reliability in each category.

Many weightings and modes of payment can be introduced on the basis of this table. If the stated maximum load is exceeded, the basic slot price is increased by 1 per cent for every additional 100 tons.

For trains running empty, the basic slot price for price classes P1 to P3 is reduced by 10 per cent and that for price classes P4 to P7 by 5 per cent. For engines running light, the basic slot price is reduced by 20 per cent.

Every regular slot ordered must be paid for irrespective of its utilisation. It is possible to reserve optional slots. A reservation fee of 20 per cent of the slot price is then collected. This fee is non-returnable.

Customers who order many slots are granted a price reduction which depends on the annual total of train-kilometres.

---

1. The pricing of the various transport sector modes should also take into account the environmental social costs and propose coherent pricing rules. This presentation does not expand on this point, which would require special development. We note that in this respect it is necessary to distinguish between two independent types of questions. On the one hand, rail pricing should be considered in relation to the social costs actually taken into account in the pricing of the principal competing mode, namely, road transport. This concern for coherence might even lead to the legitimising of intermodal balancing subsidies, theoretically justified by a second-order optimum. On the other hand, and this is a totally different problem, it is necessary to introduce differential pricing within the rail mode in order to take into account the advantages and disadvantages of the different technologies used by the operators and gradually encourage the use of those that are less polluting.

2. Brooks, M. and Button, K. propose a detailed typology of these various experiences in "Separating transport track from operations: a typology of international experiences", *International Journal of Transport Economics*, Vol. XXII, No. 3, October 1995.

3. This is the case in the United States for certain passenger services for which Amtrak must negotiate an infrastructure access charge with the integrated private operators.

4. The calculation methods used in these studies can often be reduced to very simple procedures, especially as the quality of data needed for more sophisticated calculations is very difficult to obtain. Nevertheless, the statistical analysis can be made more precise by greater refinement and by establishing precise relations between particular types of costs, networks and users.

5. Avoidable fixed costs are then defined as those which would disappear if the firm stopped producing one of its products.

6. Winston, in a survey of road pricing, explains that the proposal to consider optimal pricing and optimal investment in parallel harks back to the work of Herbert Mohring (1962).

This school includes such authors as T.E. Keeler, K.A. Small, M. Kraus, S. Glaister and S.A. Morrison.

Morrison, S.A. (1986), "A Survey of Road Pricing", *Transportation Research*, Vol. 20A, No. 2, March, pp. 85-97.

Mohring, H. and Harwitz, M.I. (1962), *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston, IL.

Small, Kenneth A., Clifford Winston and Carol A. Evans (1989), *Road Work: A new Highway Pricing and Investment Policy*, Washington, D.C., The Brookings Institution.

Hau, Timothy D. (1992), *Economic Fundamentals of Road Pricing: A Diagrammatic Analysis*, Transport Division, Infrastructure and Urban Development Department, The World Bank, December.

7. Allais, Maurice (1948), *Revue d'Économie Politique*, *op. cit.*, p. 230.

8. Boiteux, Marcel (1994), "Le développement de l'approche économique du service public", *L'Europe à l'épreuve de l'intérêt général*, Collection ISUPE, Édition ASPE EUROPE, Paris, p. 46.

9. Kolm, Serge-Christophe (1968), "La théorie économique générale de l'encombrement", *Futuribles*, SEDEIS.

10. Although traffic management is generally based on rules of priority, which distinguishes the rail problem from that of the roads and creates a further difficulty for the theoretical approach to the pricing of rail infrastructure.

11. Here we have used the demonstration proposed by Steven A. Morrison, who bases himself on the work of authors such as Mohring, Harwitz and Vickrey.

12. Hotelling, H. (1938), "The General Welfare in relation to problems of Taxation and of Railways and Utility Rates", *Econometrica*, 6(3), July, pp. 242-269.

13. Kolm, Serge-Christophe (1971), *La valeur publique, l'État et le système des prix,* CNRS, Dunod, Paris, p. 7.

14. In the literature, the French school is associated with the rejection of pricing based solely on balancing the accounts. This takes little account of the analyses of Jules Dupuit in the last century or of the more recent work of Maurice Allais and Marcel Boiteux.

15. Blaug, Mark (1986), "La pensée économique, origine et développement", *Economica*, 4th Edition, Paris, p. 718.

16. Allais, Maurice (1948), "Le problème de la coordination des transports et la théorie économique", *Revue d'Économie politique*, 58, pp. 212-271.

17. Vickrey, W. (1948), "Some Objections to Marginal Cost Pricing", in

accountability of the managers. In no circumstances should this second factor be neglected." *Note Prévision*, Commissariat Général du Plan, p. 151.

18. Bernard, A. (1983), "Coût marginal ou coût moyen ?", in Quinet, E. (editor), *Les transports et la puissance publique*, Presses de l'École Nationale des Ponts et Chaussées, report on transport economics seminar organised in 1981-82 by the École Nationale des Ponts et Chaussées, p. 36.

19. Very many commentaries on this type of pricing point out that fairness implies that the consumers should bear the costs of producing the goods they consume and that all the consumers should pay the same unit price for the same good. This is a frequently recurring complaint: "those systems which differentiate between deficit tolls according to the characteristics of the demand are generally considered unreasonable and unfair". Oort, C.J., *op. cit*., p. 62.

Clearly, the theoretical considerations advanced here shatter this principle. The tolls applied may vary for products that are identical both from the technical standpoint and by reason of their cost.

20. Dupuit, Jules (1849), "De l'influence des péages sur l'utilité des voies de communication", *Annales des ponts et chaussées*, p. 248.

21. Boiteux, Marcel, "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire", *Econometrica*, No. 24, 1956, pp. 22-40.

22. Ramsey, F. (1927), "A contribution to the theory of taxation", *Economic Journal*, Vol. 37, No. 1.

24. With that "s" Maurice Allais distances himself from the standard model. For further details, see: Allais, Maurice (1971), "Les théories de l'équilibre économique général et de l'efficacité maximale. Impasses récentes et nouvelles perspectives", *Revue d'Économie Politique*, No. 3, May-June, pp. 331-409.

25. Dupuit, Jules (1849), "De l'influence des péages sur l'utilité des voies de communication", *Annales des ponts et chaussées*, pp. 170-248.

26. We shall not describe the reforms themselves as they have already been extensively analysed. See, for example:

- Montagu (1995), "L'accès à l'infrastructure ferroviaire, l'expérience britannique", *Rail International*, 26, January, pp. 5-15. Paper read at the AICCF/CCFE Conference, Berlin, 6-7 October 1994: The Railways and European Transport Policy.
- (1994), *Britain's Railways: a new Era*, The Department of Transport, London, March.
- Preston, John (1996), "The economics of British Rail Privatisation: an assessment", *Transport Review*, Vol. 16, No. 1, pp. 1-21.
- British Ministry of Transport (1996), *Transport Report*.

  Available on the Web at:

  <URL http://www.open.gov.uk/dot/ann_rpt/chap_1.htm>.
- Nash, Christopher (1996), "The British Experience", *The Separation of Operations from Infrastructure in the Provision of Railway Services*, Round Table 103, ECMT, pp. 59-102.
- Bradshaw, B. and L. Mason (1994), *Rail Privatisation: Facts, Issues and Opportunities*, Oxford Economic Research Associates Ltd., Oxford.

27. Railtrack derives its income from user fees paid by the operators (supply of electricity, etc.), rents paid for the use of stations and depots and rents from its commercial assets. To these should be added the access charges which are determined by negotiation (see below). The procedures have been progressively refined. At the beginning, no rules for calculating the charges were laid down. The first charges were fixed at a level that would cover the total costs and ensure a return on the invested capital of the order of 8 per cent.

28. The charge includes a fixed annual fee comprising the allocated fixed costs (joint costs) and the additional fixed costs (specific to each company). The fixed charges, which correspond to about three-quarters of the infrastructure costs, form the subject of negotiations between the operators and Railtrack. The variable charges contain infrastructure user fees calculated in terms of train-miles which are different for each category of rolling stock (10 per cent of total costs). The costs incurred at regional and national levels are shared out among all the franchise holders in proportion to their receipts from fares. The costs incurred at local level, or on a single line, must be distributed among the users in proportion to the number of vehicle-kilometres travelled.

29. Avoidable costs: rule for the allocation of the fixed costs of the whole of the services provided by an operator, equal to the amount saved in the event of his services being eliminated.

30. Additional costs: increase in infrastructure costs imposed by its services, taking into account the configuration of the other services.

31. The corresponding subsidies are financed from revenue generated by the petroleum tax *(Mineralölsteuer).* Note that article 4 of the Railways Restructuring Act states that, from 1996, the DB AG will no longer receive any direct funds from the Federal Government for managing regional passenger services. The subsidies are allocated to the *Länder,* which use them in accordance with their own regional transport policy. However, the *Länder* must use these transfers for public transport purposes.

Table 1 . **Quality of service reference table**

Price class	P1 High-speed traffic	P2 Express passenger traffic (main lines)	P3 Regional express passenger traffic	P4 Average-speed main-line passenger traffic	P5 Short-haul regional passenger traffic	P6 Local passenger traffic	P7 S-Bahn (urban rail transport)
Max. load (t)	1 000	750	550	750	400	400	450
Permissible speed on at least one section	200 km/h or more	Up to 200 km/h	Up to 160 km/h	Up to 140 km/h			
Planning quality	105 %	108 %	110 %	120 %	120 %	120 %	108 %
Category of train	Intercity express (ICE)	EuroCity and InterCity	InterRegio, main-line express trains	Express night trains, accompanied-car trains, fair trains	Regional express train, through train	Regional train, CityBahn, slow train	S-Bahn train

Table 2 **Percentage reduction in terms of train-km per year**

Percent reduction	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Main-line traffic train-km (millions)	14	28	42	56	70	84	98	112	126	140	154	168	182	196	210	224	238	252	266	280
Short-haul traffic train-km (millions)	0.3	3	25	63	134	205	250	293	333	370	407	444	481	518	555	592	629	666	703	740

Customers who order slots for several years and sign a contract are granted a further price reduction in addition to that mentioned above. For firm orders extending over 2, 3, 4 and 5 years, the corresponding reductions are 2, 3, 4 and 6 per cent.

The German pricing policy makes the financing of rail activities truly transparent, even though the determination of the costs in terms of train-km is far from receiving unanimous approval and constitutes an obstacle to the entry of new operators. The relatively high prices and the choice of pricing applied to the train rather than the wagon are dissuading new operators from moving in.

These price scales have introduced a certain flexibility, but it is still insufficient and, in a way, is institutionalising the status quo by discouraging the adoption of new techniques and limiting the freedom of manoeuvre of possible new entrants. Separation has not progressed very far because DB AG is still both service provider and network operator. The transition is a gentle one. The undertaking seems to have been genuinely successful since already more than sixty transport operators have moved onto the DB AG's rail network and their number is steadily increasing

While proposing rather high marginal network entry costs, this pricing system reduces, on the one hand, the uncertainty for future operators by encouraging longer-term commitments and a search for financial and technical partners and, on the other hand, short-term practices which could have pernicious effects on the continuity and quality of the rail service.

Finally, the choice of a high level of global pricing guarantees the infrastructure manager a development capability, which is one of the strong political choices of these reforms, together with the choice of a user tariff likely to lead to the optimal utilisation of the network.

4.3 The French example: a transitional phase[32]

Introduced against a background of social strife, in particular a big strike in 1995, the French reforms consisted in establishing *Réseau Ferré de France* (RFF), a new public company which, as balance-sheet liabilities, received three-quarters of the debt of the SNCF and, on the asset side, the national network infrastructure, with the exception of the stations and installations needed by the historical operator. The latter was entrusted with the management and maintenance of the infrastructure on behalf of the RFF, which pays the bill for this service (16.8 billion francs for 1997, the first year of implementation). The SNCF pays the RFF for the use of the infrastructure. For the first two years, a limit was placed on this fee (slightly under 6 billion in 1997) by the law and the decree establishing the new system.

The first characteristic of the system relates to the fact that, relieved of most of its debt and infrastructure financing, the SNCF is in a position to balance its accounts, which it is expected to do in 1999. Secondly, the new infrastructure company, which at present can only count on earning 6 billion francs, must cover, in addition to nearly 17 billion in network maintenance and management costs and unavoidable investment costs amounting to about 13.6 billion, charges of around 9 billion on the debt inherited from the reforms. Obviously, most of the difference between expenditure and income is covered by the Government, in the form of either a capital grant or subsidies, the difference being made up by borrowing.

This, then, is a very special situation which can be interpreted in two different ways. Either the RFF may be regarded as a body whose principal function is to take over the debt and cover the deficit (net of subsidies) of the infrastructure account by borrowing. Naturally, in this case, investment would be the adjustment variable and would inevitably face historical decline. Or the present situation may be regarded as a transitional phase for putting new structures in place, after which pricing that offers greater incentives and ensures better coverage of the costs will be gradually introduced.

Being capped in 1997 and 1998, the present fee system is obviously far removed from the principle of covering the costs. It corresponds to about one-fifth (in terms of the total amount) of the German system. There is little connection between the six billion constraint imposed and the reality of the actual costs, particularly as more than half of this sum comes from the regional organising authorities (mainly "Parisian" passenger transport) and less than half from the SNCF.

There is little point in studying this provisional system, precisely because it is capped. However, it should be noted that the idea was to create incentives, especially where the demand for slots is high relative to capacity, i.e. on the urban and suburban lines (the part of the network designated R_0) and to a lesser extent on the busy high-speed lines (R_1). On the other hand, on the low-density, high-speed network and on the main-line network (R_2) the fees are very low, while on the rest of the network (R_3) they are symbolic.

The fee system distinguishes between a monthly access charge, AC, per kilometre of lines for which access is requested, a reservation charge, RC, per kilometre and per slot reserved and a traffic charge, TC, per train-kilometre. There are different reservation charges for peak periods, normal periods and slack periods. The corresponding charges for 1997 are shown in Table 3.

Clearly, this system is much less detailed and sophisticated than that introduced on the German network and thus raises the question of whether it is sufficiently refined to enable the relevant marginal costs and homogeneous demand segments to be distinguished in the event of the future system being steered towards a more determinedly economic form of pricing.

Table 3. **French provisional pricing**

(1997 unit prices in francs)

Sub-network	R_0	R_1	R_2	R_3
AC	11 000	11 000	250	0
RC (peak)	100	18	0.85	0
RC (normal)	44	6	0.85	0
RC (slack)	20	4	0.85	0
TC	0.30	0.30	0.30	0.30

Note: AC is expressed per month and per km of line, RC per slot-km, TC per train-km.

In the French case, clearly the main problem is how pricing will evolve after 1999. This question is overshadowed by the fact that, overall, the rail system is running at a loss. The first step then will be to choose between two strategic directions. One choice would be a low-toll system which would concentrate the public contribution on covering the deficits of the RFF and financing new investment. In this case a policy of long-term marginal cost pricing without budgetary constraint might be envisaged. A second choice would be a system that combined budgetary constraint with a Ramsey-Boiteux principle. In this case the SNCF would have to be subsidised for a fairly long time, but the subsidies could be correlated with the loss-making services thus financed, thereby allowing the authorities latitude to compare their cost and their utility.

On the basis of a study in progress, the RFF is to propose to the Government a user pricing system designed to encourage a better allocation of resources.

5. CONCLUSION

To reach a conclusion on such a subject would be to suppose that a definitive theoretical contribution, which was both coherent and pertinent and proposed measurable concepts, would make it possible to solve, down to a few details, this difficult rail infrastructure pricing problem. Only a patient approach that takes into account all the attempts to apply theoretical prescriptions will enable us to work towards a satisfactory solution.

It is no insult to the achievements of economic theory or railway economics to conclude with the following few lines which were written about a century and a half ago:

We merely wished to show that the way in which the tolls are fixed can greatly extend the utility of certain routes and that the guiding principle in assessing these charges should not be to set a price proportional to the weight or the distance nor to favour a particular industry or a particular class of passengers, but rather to impose on each passenger and on each good only a price that is lower than that which would prevent the passenger or good from using the route. Admittedly, the methodical classification of these passengers and goods does call for inventiveness and an intimate knowledge of the local circumstances, but a sound theory can do much to facilitate this work[33].

NOTES

32. See the following official texts:

- Law No. 97-135 of 13 February 1997, establishing the public corporation, *Réseau Ferré de France*, with a view to the revival of rail transport;
- Decree No. 97-446 of 5 May 1997 on national rail network user fees;
- and, finally, the Orders of 30 December 1997 on national rail network user fees, JO, 31 December 1997, pp. 19461-19463.

33. Dupuit, Jules (1873), "Poids et mesure", in Coquelin and Guillaumin (eds.), *Dictionnaire de l'économie politique*, 4th Edition, Vol. 2, Paris, pp. 339-344.

BIBLIOGRAPHY

Allais, Maurice (1945), *Économie pure et rendement social*, Edition Sirey, Paris, 72 pp.

Allais, Maurice (1948), "Le problème de la coordination des transports et la théorie économique", *Revue d'Économie politique*, 58, pp. 212-271.

Allais, Maurice *et al.* (1965), Options de la politique tarifaire dans les transports, Report prepared at the request of Commission of the European Economic Community, *Studies Coll., Transport Series, No. 1*, Brussels.

Allais, Maurice (1989), "L'économie des infrastructures de transport et les fondements du calcul économique", *Revue d'Économie Politique*, 99, No. 2, pp. 159-197.

Allais, Maurice (1989), *Théorie générale des surplus*, PUG, Presse Universitaire de Grenoble (2nd edition), 716 pp.

Banister, David, Kenneth Button *et al.* (1993), *Transport, the Environment and Sustainable Development,* Alexandrine Press, Oxford, 275 pp.

Baumstark, L. (1997), "Tarification de l'usage des infrastructures et théorie de l'allocation optimale des ressources, de la logique des coûts à la révélation des préférences", Doctoral thesis, Université Lumière Lyons 2.

Boiteux, Marcel (1956), "La vente au coût marginal", *Revue Française de l'Énergie*, December.

Boiteux, Marcel (1956), "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire", *Econometrica*, No. 24, pp. 22-40.

Bowers, Ph. (1995), "Commercialisation, accès local et international aux réseaux ferrés", *Public Transport International*, No. 1.

Brooks, M., K. Button (1995), "Separating transport track from operations: a typology of international experiences", *International Journal of Transport Economics*, Vol. XXII, No. 3, October.

Council Directive 91/440 of 29 July 1991 on the development of the Community's railways, OJ L 237 of 24/08/1991, 25 pp.

Council Directive 95/18/EC of 19 June 1995 on the licensing of railway undertakings (OJ L 143/70-74).

Council Directive 95/19/EC of 19 June 1995 on the allocation of railway infrastructure capacity and the charging of infrastructure fees (OJ L 143/75-78), 27 June 1995.

Regulation (EEC) No. 1191/69 of the Council of 28 June 1969 (OJ L 156) on action by Member States concerning the obligations inherent in the concept of a public service in transport by rail, road and inland waterway, last amended by Regulation No. 1893/91 of the Council, OJ L 169 of 29/6/1991.

Regulation (EEC) 1192/69 of the Council of 26 June 1969 on common rules for the normalisation of the accounts of railway undertakings, OJ L 156 of 28/06/1969.

Regulation (EEC) No. 1107/70 of the Council of 4 June 1970 on the granting of aids for transport by rail, road and inland waterway, OJ L 130 of 15/06/1970.

Green Paper (1995), *Towards fair and efficient pricing in transport,* COM(95)691 of 20/12/1995, 58 pp.

White Paper (1996), *A strategy for revitalising the Community's railways*, COM(96) 421 final, Brussels, 30 July 1996, 54 pp.

ECMT (1995), *Why Do We Need Railways?,* Proceedings of International Seminar, Paris, 19-20 January, 226 pp.

ECMT (1993), *Privatisation of Railways*, *Round Table 90*, Paris.

ECMT (1997), *The Separation of Operations from Infrastructure in the Provision of Railway Services*, *Round Table 103*, Paris.

Commissariat Général du Plan (1995), *Transport, le prix d'une stratégie* (Workshop chaired by A. Bonnafous), Vols. 1 and 2, La Documentation Française, Paris.

Commissariat Général du Plan (1996), *Transformation des opérateurs publics de réseaux en Allemagne -- Les cas du contrôle aérien, des transports ferroviaires, de la poste et des télécommunications.* Study made by Cabinet ESL & Network, Vols. 1 and 2, March.

Commissariat Général du Plan (1993), *Transports : pour une cohérence stratégique*, Workshop on the strategic directions of transport policy and their medium-term implications, Commissariat Général du Plan, September.

Commissariat Général du Plan (1994), *Transports : pour un meilleur choix des investissements* (Group chaired by Marcel Boiteux), La Documentation Française, November, 132 pp.

Commissariat Général du Plan (1995), *Transports : le prix d'une stratégie* (Workshop chaired by Alain Bonnafous), Vols. 1 and 2, La Documentation Française, pp. 213 and 182.

DB AG (1996), "Der Trassenpreiskatalog der DB AG: ein Vorbild für Europa?" Lecture given by Mrs Haase at the 29th Scientific Transport Research Seminar, "Competition Policy in Deregulated Transport Markets", organised by the University of Freiburg (Germany), Hinterzarten, October.

DB AG, *Trassenpreise - Güterverkehr*, Frankfurt am Main, 1995.

DB AG, *Trassenpreise - Personenverkehr*, Frankfurt am Main, 1995.

Dodgson, J. (1994), "Railway privatisation", in: M. Bishop, J. Kay and C. Mayer (eds.), *Privatisation and Economic Performance*, Oxford University Press, Oxford.

Drèze, H. (1994), "Marginal-Cost Pricing", Introduction to Part III in: Arnott, Richard, Kenneth Arrow, Anthony B. Atkinson and Jacques Dreze (eds.), *Public Economics* (Selected papers by William Vickrey), Cambridge University Press, pp. 190-195.

Ewers, H.-J. (1993), Privatisierung und Deregulierung bei den Eisenbahnen: das Beispiel der DB und der DR, Institut für Verkehrswissenschaft, Universität Münster, 15-17 September, 11 pp.

Gillen, David W. and Tae Hoon Oum (1991), *Transportation infrastructure: Pricing, Investment and Cost Recovery,* Royal Commission on National Passenger Transportation, Technical Seminar, Ottawa, Canada, September 26 and 27, 181 pp.

Guihéry, L. (1996), "Fédéralisme allemand et réforme des chemins de fer en Allemagne : le transport ferroviaire régional de voyageurs devient une nouvelle compétence des Länder", *Documents-revue des questions allemandes*, No. 1, Paris, p. 79.

Guihéry, L. (1997), "Réforme des chemins de fer et tarification d'usage des infrastructures ferroviaires en Allemagne, une organisation contrôlée de la concurrence sur le réseau ferré allemand", *Transport,* No. 382, March-April, pp. 112-121.

Hau, Timothy D. (1993), "Economic Fundamentals of Road Pricing: A Diagrammatic Analysis", Working Paper, WPS 1070, Transport Division, Infrastructure and Urban Development Department, The World Bank, p. 87.

Horton, Sir Robert (1997), "Re-engineering Britain's railways: risks and rewards", *European Railway Review*, June, pp. 89-95.

Hotelling, H. (1938), "The general welfare in relation to problems of taxation and of railways and utility rates", *Econometrica*, 6(3), July, pp. 242-69.

Hutter, R. (1950), "La théorie économique et la gestion commerciale du chemin de fer", *Revue Générale des Chemins de Fer*, February, July and October, p. 44.

Kolm, Serge-Christophe (1968), "La théorie économique générale de l'encombrement", *Futuribles*, SEDEIS, p. 82.

Laffont, Jean-Jacques (1977), *Effets externes et théorie économique*, Monographies du séminaire d'économétrie, No. 13, éditions du CNRS, Paris, 200 pp.

Laffont, Jean-Jacques and Jean Tirole (1994), *A Theory of Incentives in Procurement and Regulation,* The MIT Press, Cambridge, Massachusetts, London, 705 pp.

Le Duc, M. (1995), *Services publics de réseau et Europe : Les exemples de l'Allemagne, de l'Espagne, de l'Italie et du Royaume-Uni*, Les études de La Documentation Française, Paris, 1995.

Morrison, Steven, *The Theory of Optimal Highway Pricing and Investment: Comment*, Northeastern University, Boston, Massachusetts.

Nash, C. (1993), "Developments in transport policy: rail privatisation in Britain", *Journal of Transport Economics and Policy*, Vol. XXVII, No. 3, September, pp. 317-322.

Office of the Rail Regulator, *Framework for the approval of Railtrack's track access charges for freight services*, a consultation document, October 1994, 19 pp.

Preston, J. (1996), "The economics of British Rail privatisation: an assessment", in *Transport Reviews*, Vol. 16, No. 1, pp. 1-21.

Quinet, E. (1990), *Analyse économique des transports*, PUF, Collection Économie, Paris, 302 pp.

Allais, Maurice, Del Viscovo, Duquesne de la Vinelle, Oort and Seidenfus (1965), "Options de la politique tarifaire dans les transports". Report prepared at the request of the Commission of the European Economic Community, *Studies Collection, Transport Series*, No. 1, Brussels.

Rathery, A. (1984), "De nouvelles perspectives pour la politique des infrastructures de transport dans les Communautés européennes ? Possibilités offertes par une approche pragmatique de la tarification d'usage des infrastructures", *Transports*, No. 297, October, pp. 358-365.

Rothengatter, W. (1991), "Deregulating the European Railway Industry: theoretical background and practical consequences", *Transportation Research*, Vol. 25A, No. 4, pp. 181-191.

Small, Kenneth A. (1990), "Highway Infrastructure: Crisis in Finance or Crisis in Management", *Review of the University of California Institute of Transportation Studies*, Vol. 13, No. 2, February, pp. 4-8.

Small, Kenneth A., Clifford Winston and Carol A. Evans (1989), *Road Work: A New Highway Pricing and Investment Policy*, The Brookings Institution, Washington, D.C.

The Department of Transport (1993), *Rail Freight Privatisation (the Government's proposals)*, London, 27 pp.

Vickrey, W. (1987), "Marginal- and Average-Cost Pricing", in: Eatwell *et al.* (eds.), *The New Palgrave Dictionary*, Vol. III, Macmillan, pp. 311-318.

Vickrey, W. (1994), *Public Economics* (Selected papers by William Vickrey), Edited by Richard Arnott, Kenneth Arrow, Anthony B. Atkinson and Jacques Dreze, Cambridge University Press, 558 pp.

Walters, A.A. (1987), "Congestion", *The New Palgrave Dictionary*, pp. 570-573.

Wilkinson, A.W. (1992), "La situation politique des British Railways", *Le Rail*, July, No. 33, p. 28-2.

SUMMARY

Paris/Lyons, January 1998

1. INTRODUCTION

The rail sector very quickly came to be regarded by economists as a typical example of a "natural monopoly". In fact, like other networks, though often even more emphatically, it displays all the characteristics which tend to compromise the theoretical pact between the mechanisms of the competitive market and the optimum allocation of resources (in the Pareto sense). On the one hand, on such a market the break-even prices lack the qualities necessary to induce the economic agents to take optimal decisions. On the other hand, it is not possible to find a satisfactory price system that enables an optimum to be decentralised on the basis of the decisions of individuals.

The sector is, in fact, characterised by increasing returns resulting from multiple indivisibilities: functional indivisibilities such as the co-ordination of activities between upstream and downstream of production, for example, between infrastructure choices and commercial policy, but also technical indivisibilities such as that determined by the continuity of the network. To this there should be added various effects such as "economies of scope" which enable the operator, by diversifying the services offered, to reduce unit costs or important network externalities, both positive and negative, or, finally, investment irreversibilities. Thus, the rail sector accumulates the conditions for the appearance of what are commonly called *market failures*. This largely explains the sector's mode of operation in the early 80s and the major role played by government throughout its history.

However, the efficiency of this monopolistic organisation and the frequent public intervention in the sector have gradually been called into question on two counts. Firstly, it has been noted that the theoretical justifications for such an organisation apply only to part of the monopoly (mainly the infrastructure). Secondly, the very real advantages afforded by such an integrated structure have been offset by its no less real disadvantages (mainly its inability to exert sufficient pressure on production costs). Thus, the policies applied in this sector have sought to eliminate these weaknesses by reintroducing, wherever possible, various kinds of competitive pressure. Infrastructure access charges have become a keystone of the reforms. This question forms the subject of the first part of our paper.

There have been many modern theoretical developments in this field and our presentation does not seek to be exhaustive in this respect. However, there are two theoretical principles that deserve closer examination and these will be the focus of our attention in the second part[34].

The first principle states that infrastructure pricing cannot be separated from investment choices. The setting of "optimal" user fees presupposes the prior definition of a certain quality of service. For example, from the quality of service required there follows an acceptable saturation level which, in its turn, determines a co-ordinated policy of investment and demand management through infrastructure user fees. Hence, the theoretical economic toll is a composite toll, comprising a cost toll relating to the costs which can be directly allocated to the users of the network and an adjustment, sometimes called a pure toll, which enables the manager to produce the necessary quality of service by adjusting, as precisely as possible, the available capacity to the expressed demand (note that this second component only concerns the part of the network with no surplus capacity).

The second principle involves relating pricing to the difficult question of covering the fixed and joint costs which within the rail system, as in many other networks, represent a large proportion of total costs. The traditional theoretical solution based on marginal cost pricing, considered optimal, was to have these costs covered by the State budget. However, the perverse effects of systematic subsidisation, nowadays decried as public failures, mean that such a solution can no longer reasonably be regarded as optimal. Nevertheless, imposing budgetary constraint on the enterprise responsible for the infrastructure is not sufficient in itself to ensure the efficiency of the system. The pricing method and the formulas for allocating the fixed costs which, by definition, cannot rationally be allocated to any particular user, have definite implications for the social surplus which the system can produce. One of the main difficulties encountered in connection with pricing is that associated with the manager's technical and economic (but also political) ability to charge differential, even discriminatory fees.

In the last part, we compare these abstract principles with actual practice.

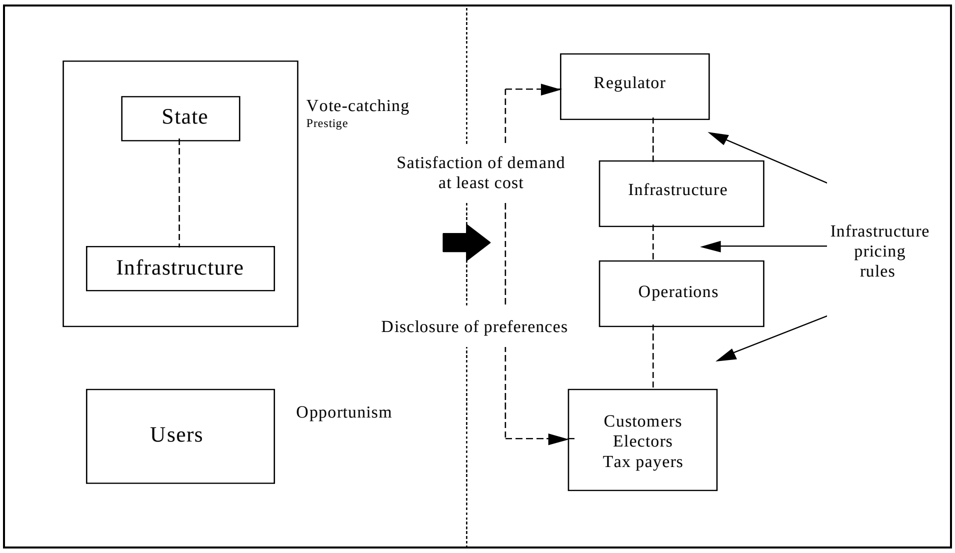
2. INFRASTRUCTURE PRICING, KEYSTONE OF INSTITUTIONAL INNOVATION

2.1 Network opening, redistribution of roles and pricing system

The idea of opening up the network is not a new one. Thus, networks have been identified in many different industries: the high-tension lines in the power distribution sector, optical fibre systems in telecommunications and delivery and collection in the postal sector, to which could be added numerous services such as the transport of electricity and information and the routing of mail. Although, as far as the network is concerned, it seems preferable to retain a monopolistic structure because of the economic characteristics of the infrastructure, this is no longer true of the services for which, on the contrary, competition seems possible.

In order to solve the problem of opening up the network, it is necessary to visualise how certain decisions which were taken internally within a single institution can be made the subject of commercial negotiations. The analysis must therefore focus on the types of relations that may exist between the decisionmakers, the private or public operators and the users. The configuration of the system and the actors is illustrated by the diagram reproduced in Figure 1.

Figure 1. **A traditional organisation chart**



It is thought that, under certain conditions, the intrusion of economic mechanisms into the integrated model will introduce transparent processes, thus obliging the various actors to disclose their preferences. This then leads to the explosion of an extreme situation in which the State produces the infrastructure regardless (or almost regardless) of the demand, of which it has only a faint idea. In this situation, the user benefits from a good without knowing its cost.

This can be improved upon by identifying intermediate levels. It is possible to distinguish between the production of the infrastructure and its management. Attention may also be focused on the behaviour of the agents and its determinants: the efficiency of the system as a whole depends on the efficiency of the means deployed to encourage and promote partnership between the economic actors and the authorities at various administrative levels. In short, it is a question of organising effective modes of interaction between the public and the commercial spheres.

On the one hand, it is necessary to establish relations between the State or, more generally, the authorities responsible for organising public transport and those directly concerned with the infrastructure so as to minimise infrastructure costs. This is part of the new theory of regulation which seeks to define more precisely the contractual framework within which these relations are to fit.

On the other hand, it is necessary to make sure that all the transport operators have access to the infrastructure on fair terms. This poses a specific problem with respect to the relations to be established between the infrastructure operator, public or private and the downstream service providers. Should or should not the monopoly be authorised to supply the market downstream? Can the sole and final responsibility for the allocation of timetable slots be entrusted to a body which itself also operates the infrastructure?

The determination of the prices at which the infrastructure manager opens up his network is one of the major considerations for the sector as a whole. These access charges are important since they determine the terms of competition between modes of transport. They must also ensure the overall efficiency of the system, that is, be sufficiently high for the upstream monopoly to be able to offer a suitably maintained infrastructure where the rail system is pertinent, but not so high as to bar the arrival of new entrants. More generally, pricing policy makes it possible, on the one hand, to encourage users to programme their services better (in the sense of making the best possible use of capacities) and, on the other, to steer investment towards ensuring that the system is developed and, in particular, the network expanded, so as to give the best possible cost-benefit ratio.

When he wishes to formalise the importance of the pricing system in the resource allocation process, the economist generally analyses the consequences of price distortion. Whatever the reasons, legitimate or not, for the distortion, the consequent loss of global surplus needs to be estimated. Within this theoretical framework, assuming an initial situation in which the pricing system is optimal, any change in prices not justified by a change in the cost of the production factors or by a sudden imbalance between supply and demand would inevitably lead to the misallocation of resources, a poor investment policy and, ultimately, less user satisfaction. In this case, the surplus available in the system would no longer be maximised.

This theoretical approach, even though it clearly relies on mechanisms less complex than they are in reality, provides valuable benchmarks for the orientation of supply and ensures cohesion between different strategies capable of serving the same purpose. The pricing system may then be understood as an information and co-ordination system, and if it offers a certain number of minimum qualities, the regulation of the system, being based to a greater extent on the rationality of the economic agents, will gain in efficiency.

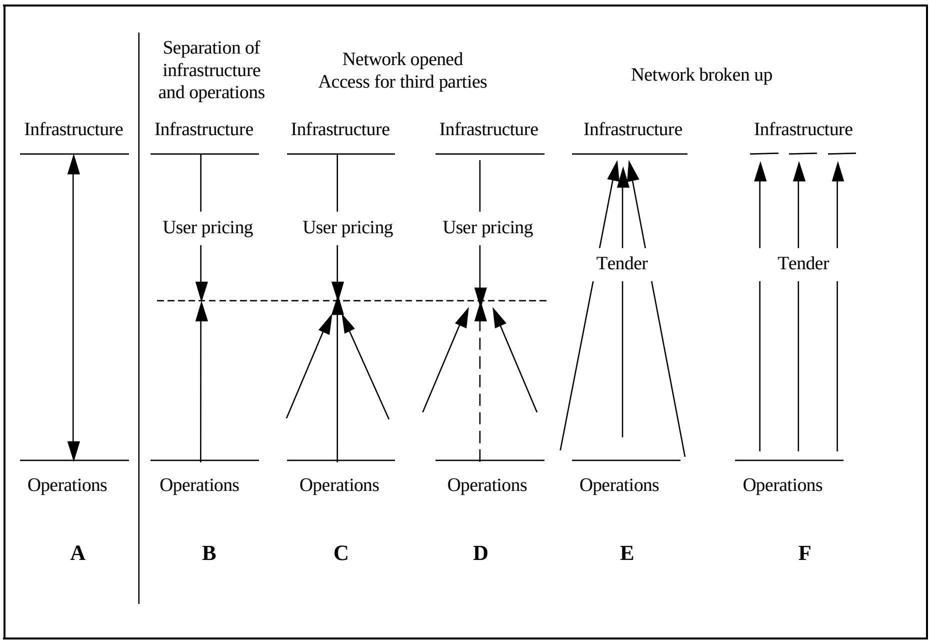
2.2. Degree of vertical separation in the rail sector and the question of access charges and user fees

These theoretical analyses underlie many of the recent trends in the European rail sector. They form the context for the adoption of European Directive 91/440, which first provided the impetus for the separation of infrastructure from operations. The Council of Ministers took a further step by issuing two new Directives 95/18 and 95/19. Before the network can be opened up, the conditions of network access must be established: licensing system, procedures for the allocation of existing capacities, coverage of infrastructure costs, network development, appeal procedures, etc. However, this Community drive is based on the reforms which many countries have already started to introduce. Although these reforms may employ different means, they apply the same basic principles, namely, the principles of decentralisation and the organisation of competitive pressure.

In practice, these widely introduced vertical separation policies are very diverse. This diversity contrasts sharply with a very homogeneous theoretical discussion of the question35. The main difference between all these systems appears to relate to the role of the State. Sometimes, though this is not the most frequent case, the State more or less organises the service, whereas the infrastructure is private36; in another configuration, the service may form the subject of a private monopoly, whereas the infrastructure remains in public hands (as in New Zealand). Under a third scenario, private companies operate a network in "partnership" with another private partner who owns the infrastructure (as in the United Kingdom).

The infrastructure user fee structure and its possible impact on the markets cannot be understood without taking into account the institutional configuration governing the management and operation of the network. A summary typology of the possible configurations is proposed in Figure 2.

Figure 2. **Typology of institutional configurations**



Situation (A) is the one that has been most frequently inherited from the past, namely, an integrated network occupying a monopoly position. As we have seen, adaptation, which may go as far as radical reform, involves creating levels of responsibility that give rise to commercial partnerships and to the introduction, with varying degrees of firmness of resolve, of the machinery of competition. Depending on the extent to which the network is broken down, several options are possible.

Assuming that the integrated monopoly is maintained, it is possible to imagine the introduction of competitive mechanisms similar to those in effect for certain urban public transport systems: invitations to tender for the management of the entire system lead to the appointment of a manager for a specified period. The integrated monopoly may be completely preserved (E). However, the network may also be broken up and tenders invited for each segment (F). The network may be divided up on a geographical or a sectoral basis, for example, with a network specialising in freight. In these various cases, the question of access charges does not really arise since the integrated monopoly is preserved: the bargaining takes place within the enterprise itself without involving any market services. The authorities may face other regulatory problems, but they will no longer have to deal directly with the question of access charges.

This does not apply to configurations (B), (C) and (D): although the integrity of the system as a whole is maintained (the infrastructure manager has authority over the entire network), the relation between the infrastructure and the operator or operators becomes distended. There are still many possibilities for reform, depending on the degree of separation between the infrastructure manager and the operator. Here the competitive pressure is exerted (with varying degrees of firmness) at the level of the infrastructure user by the (more or less) explicit means of diversification and dynamisation of the supply of services. However, in all three cases, it is a question of sending the operators a price signal calculated to influence their demand for the use of the infrastructure. The question of user fees will then be the determining factor.

Case (B) illustrates the lowest level of separation: third parties are denied access. Here, separation is intended simply as a means of clarifying the enterprise's accounts by dividing the expenditure items more clearly between those that relate to investment and maintenance costs and those that relate to operations. This model corresponds to the situation of those countries which, following the Community decision, preferred not to call into question either the integrity or the unity of their monopoly. Thus, the whole of the network continues to be run by a single operator. However, this system seems rather unstable and can be better understood as a staging post on the road to a more radical mode of organisation in which, separation having been made effective, third-party access can be encouraged. In the first case, pricing can remain a sort of accounting device by means of which the carrier is supposed to remunerate whoever undertakes to maintain and develop the infrastructure, that is to say, himself. Thus, the pricing rules are not really market mechanisms within which the parties endeavour to optimise their results. In cases (C) and (D), which are of more particular interest, the situation is quite different.

In case (C), the opening up to third parties is explicit but there remains a principal operator. Thus, the opening is still marginal and may be restricted to certain market segments, for example, the international market. This transitional situation may evolve into the much more competitive situation (D) in which, third-party access having been generalised, the historical operator is one actor among others. These two models lie at the heart of the two main scenarios for the implementation of the reforms undertaken in Europe.

We note, however, that the choice of system (C) does not protect the network concerned from sliding towards system (D). There will be de facto involvement of the international services and their demand for track will have to be reconciled with that of the national operator. Here, Community policy has clearly taken the same path as led road freight transport from progressive liberalisation of international transport to complete freedom of cabotage.

Consequently, the opening up of the network, however partial, raises the problem of how to regulate between the operator already in place and other potential operators. Even though the slot allocation rules may play a very important part, in the long run the calculation of access charges should prove decisive. Suppose, for example, that the allocation rules favoured the operator in place, say, through a practice involving "grandfather clauses" comparable to that observed in air transport, but that, at the same time, there was a heavy demand for slots, for international traffic, for example. The infrastructure monopoly could then take advantage of this excess of demand over available capacity by increasing the user price where the supply was scarce. If the use of the "grandfather clause" were too expensive for the historical operator, he could waive it and review his operating schedules in order to fall back on slots in better supply and the available capacity would then be sufficient. If it were still profitable to exercise the right despite the higher price, this would bring the infrastructure monopoly a return sufficient to finance the necessary investment in capacity.

In practice, this configuration might be relatively rare, but a similar problem might arise, including in the case of configuration (B), if the historical operator were to make, within the same space-time frame, requests for slots on behalf of different services (national and regional passenger services, freight services, etc.). If these requests were sufficiently well justified because of corresponding traffic demand, that would reduce to the expression of a need for investment in capacity on the sections in question. The historical operator might then have an interest in sending a "price signal" reflecting the shortage of available slots to each of the requesting services. It would be up to the latter either to buy because they were able to cover the cost or to transfer the demand to less busy routes or periods. This comes back to the above-mentioned choice between managing the demand and being able to finance additional capacity.

This alternative reduces to an infrastructure pricing principle which is well known but worth recalling in order properly to explore its conditions of application to rail transport.

3. INFRASTRUCTURE PRICING AND MAXIMUM EXTRACTION OF AVAILABLE SURPLUS

Infrastructure pricing may have more than one objective. The following presentation is founded on the idea that the basic principles should be established as a function of the collective efficiency of the system. Indeed, it would seem difficult to defend a pricing system which regularly deviated from that objective. Thus, this theoretical look at the problem suggests that there should be no segmenting of pricing between different, possibly contradictory objectives, namely, optimum use of the infrastructure, on the one hand, and the financing of its renewal or the pursuit of social or environmental objectives on the other. In fact, all these objectives can be reduced to the general objective of maximising the global surplus generated by the system.

On what principles should pricing be based in order to extract the maximum surplus from the system for sharing out among the various actors? This surplus can be identified only by relating it to the costs of the system and, in particular, the investment costs.

3.1. The concept of marginal cost and the optimisation of investment

Maximum efficiency obtained by marginal cost pricing in sectors with increasing returns is a decisive conclusion contributed by the theory of welfare economics. However, this theoretical result presents a number of problems. Apart from the undoubted difficulties of application, the concept is also open to theoretical objections.

In many cases, however, in discussing this issue, confusion is created insofar as the concept of marginal cost that is criticised is very often reduced to a user cost[37]. At first glance, this is hardly surprising. Understandably, since the marginal cost enters into the equation as the derivative of a total cost function, any cost factor that does not vary with production will vanish in the mathematical operation.

This classical presentation, without actually being false, can lead to inaccuracies with dangerous theoretical and practical consequences, simply because in this approach it is assumed that the investment is given and realised. In a manner of speaking, the act of investment is ignored as an optimising tool, on the same basis as pricing. However, the theoretical framework of the optimal allocation of resources cannot be satisfied with this reduction or with the solution consisting in transposing this short-term static model to the long term and treating the fixed costs as variables, thus causing them to vanish.

Indeed, taking the fixed costs into account presupposes an analytical approach quite different from that which consists in making them vanish. The fixed costs form part of the irreducible and fundamental linkage between the short and the long term. Leaving out that linkage amounts to evading the embarrassing questions raised by coverage of the costs thus incurred over several periods. From the standpoint of cost minimisation, it is well known that it is always possible to distinguish between directly avoidable costs[38], costs that can be gradually absorbed and, finally, totally unavoidable costs. This distinction then makes it necessary to bring other concepts into play.

In particular, it is necessary to introduce a terminological distinction in order to avoid confusion. The variable marginal cost is a partial cost clearly distinguishable from the marginal cost, which has a different theoretical content: the partial cost relates specifically to a short-term situation in which only the variable costs are considered, while the concept of marginal cost relates to a situation in which it is assumed that the authorities meet the demand, if need be by carrying out a project that creates a discounted surplus greater than the total of the costs necessary to produce it. Thus, as many authors have stressed[39], this concept is fundamentally linked with the investment decision.

Taking the long term into account requires the use of a programme situated upstream of the decision to invest. The fixed costs are not, strictly speaking, variabilised; however, the transport infrastructure is regarded not as a natural resource made available but as a good, to produce which the community must sacrifice resources.

Suppose that there are several technologies capable of meeting a given level of demand. Each of these can be characterised by a fixed cost, independent of the level of utilisation of the infrastructure and by a variable cost. The producer or the authorities have a certain number of possible solutions available to meet the demand. To each investment there corresponds a particular total cost function which depends on two variables, the investment I [10] and the level of production q:

$$C(I,q) = a(I) + b(I,q) \quad {}^{11}$$

Thus it is possible to distinguish between two marginal costs:

- the short-term marginal cost when q varies for a given investment level:

$$\gamma = \frac{\partial C(I,q)}{\partial q} \quad {}^{12} \quad ;$$

- the long-term marginal cost when the investment varies: $\Gamma = \frac{\partial C(I,q)}{\partial I}$. [13]

It is reasonable to assume that the producer (or the authorities) wishes to produce the good or service in question at the lowest possible price, if only to maximise the collective surplus. Thus, for each anticipated level of production, he will adopt the technology calculated to supply that good or service at the lowest price. Therefore, to each level of production there corresponds a minimum cost which itself relates to a particular total cost function belonging to the family of $C(I_i)$[14]. If it is assumed that I can vary continuously, the totality of these optimal points, which minimise the cost of production, will form a so-called long-term cost curve. We note that the curve obtained as a result of such a theoretical projection will be completely virtual. It defines a boundary of maximum efficiency of the production system at a given instant of time. Rather than an objective to be achieved, it is primarily a theoretical benchmark. It is a short-term curve envelope. This curve defines a functional relation between the long-term total cost and the quantity produced, on condition that each level of production is obtained with plant of the optimum size.

If it is assumed that the producer is able, at any time, to adapt his investment to the level of demand, there will be an infinity of short-term total cost functions. The system may be called upon to display its adaptability at any time. In this case, there is no contradiction between the short-term and long-term marginal costs. The short-term marginal costs are always defined, while the size of the plant is optimal for the production level in question.

To discard the hypothesis of permanent adaptability (divisibility) is to accept that the demand can be met by a technology or investment which is not necessarily optimal. In this situation, which is generally that of transport systems, there is almost inevitably a difference between the short-term and long-term marginal costs. There will then be several possibilities.

In the first case, that of underinvestment, the technology employed is not optimal since the short-term marginal cost is higher than the long-term marginal cost. In the second case, that of overinvestment, the short-term marginal cost is lower than the long-term marginal cost; the demand is met by an investment which costs too much for the use which is made of it. When, exceptionally, the system is optimised, the supply is precisely adapted to the demand, the two marginal costs are equal and, in this case, the marginal cost involves, over and above the short-term marginal expenditure, the optimal variations in the cost of the plant necessary to meet the demand.

> "*Under these conditions, and for practical applications of the theory of social returns, it is necessary to define the marginal cost of a specific service as the additional costs of all kinds (labour, energy, raw materials, depreciation, interest charges) resulting from the provision of an additional unit of that service when the existing fixed plant is precisely adapted to the volume of production in question*[40]."

Accordingly, the marginal cost to which pricing theory relates is defined at economic equilibrium and, consequently, with total utilisation of the fixed plant. This leads to a radical distinction between the marginal cost as described above and the partial cost. Thus, Marcel Boiteux considers it necessary, despite everything that may have been written on this subject during the last thirty years, to denounce, once again, the frequent identifying in common parlance of the marginal with the variable cost. He considers this "faulty" identification to be still "a major source of misunderstanding"[41].

This also means that the determination of the marginal costs has real economic significance only on the assumption that the infrastructure is optimally managed. For the concept to be pertinent, the maintenance operations and the investment renewal operations must be carried out at the optimal time and their cost must be minimised.

If the marginal costs thus defined are used for pricing purposes, the system will be coherent and optimal. This means that the consumers' choices will be optimally oriented since they will be encouraged to choose, among several ways of satisfying their needs within a given framework of constraints, that which is least expensive for the community. However, at the same time, the community must ensure that the demand thus expressed is always satisfied at the least possible cost.

This clarification is a necessary preliminary to the application of practical pricing tools. Within this context it is now clear that marginal cost pricing is future-oriented pricing. Accordingly, it should concern itself not with the previous cost of the infrastructure but with the use that the users will make of it. A knowledge of the users' preferences then becomes one of the keys to the global optimisation of the system from the standpoint of both economic calculation and pricing.

However, the balance between supply and demand thus achieved may give rise to very different situations on the rail system depending on the level at which the service quality is to be pitched. We have previously used expressions such as "meeting the demand" or "plant of optimum size". For example, should there be perfectly regular timetables with zero delays or would a reasonable amount of delay be acceptable if it permitted the more intensive use of capacity? Clearly, these questions prompt us to reflect on the quality of service to be provided which, in its turn, is also subject to a trade-off between costs and benefits.

3.2 The supply-demand equation or how to define an optimal quality of service

To deal with this problem, public economics traditionally relies on the notion of a mixed collective good, i.e. a good such that the quantity consumed can be distributed among the individual consumers (the good is therefore divisible): the consumption of a good by a user cannot be consumed by another, whereas certain other so-called quality characteristics remain indivisible because they concern all users. These mixed collective goods are subject to so-called congestion effects42. In the case of rail infrastructure, the indivisibility of its use can never be ensured inasmuch as two trains cannot share the same "slot" at the same time. The slots are shared between the different users. The use of infrastructure, from this point of view, is divisible. On the other hand, the quality of service, which depends on the reliability of the train timetables on a congested part of the network, is the same for all the users at a given instant43. This collective good is called mixed because it has two fundamentally different characteristics, the first -- access to infrastructure -- being divisible, the other not. Congestion is a particular case of an external effect *where the reasons why people cause and suffer it are linked with the consumption of the same service.*

In order to optimise the system and achieve the maximum available surplus on the infrastructure, the authorities must both determine the optimal level of investment and, through good pricing, manage the level of demand. Investment and level of demand are the two factors of the quality of service offered which different users value in varying degrees.

In order to isolate the problem posed by this linkage, let us consider an infrastructure, the cost of using which is very low and may therefore be deemed negligible44.

Let us suppose that all the users i liable to use the infrastructure place different values on the use of this good, for a given quality of service. We then obtain an inverse demand curve which we will denote by $P_i(Q_i, g)$, where Q_i represents the utilisation of the infrastructure by the individual i, and g an indicator of service quality.

The quality of service offered on this infrastructure depends on two parameters: the number of users on the infrastructure and its physical characteristics, which depend on the level of investment k. This level of investment varies with time as a function of the outlay made by the operator, $\psi(k)$ 15.

Taking into consideration several time periods t, we write $P_{it}(Q_{it}, g_t)$, with $g_t = (Q_{it}, ..., Q_{nt}, k)$ the quality of service at period t. We then suppose that each user integrates this quality of service, $P_{it}(Q_{it}, g_t)$ being the value that the user i

accords to the use of the infrastructure at a given level of quality g. This value may evolve over time.

The individual surplus is given by the difference that exists between the value that the individual attributes to the use of the infrastructure and the cost which that use represents for him, remembering that in this case the user cost, in its strictest sense, is seen as negligible. This cost is therefore limited to the deterioration in quality of service suffered by the user, due to the utilisation of the infrastructure by others. This cost is all the more important the greater the number of users and the weaker the investment. If we take D_t as the deterioration of service quality at time t and V_{it} the value that the user i gives to this deterioration, the surplus of user i for each period t is expressed by:

$$P_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right)$$

Now let us assume that the operator is seeking to maximise the social surplus which this infrastructure is capable of generating. For each individual i and for the entire length of time in question, we have:

$$\sum_{t=1}^{T} \left(\int_0^{Q_{it}} P_{it} . dQ_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) \right)$$

The global surplus is obtained by summing over all the users and subtracting the investment cost. This calculation leads to the following expression:

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \left(\int_0^{Q_i} P_{it} . dQ_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) \right) - \psi(k)$$

The maximisation of the surplus depends on two variables, the utilisation of the infrastructure and the level of investment.

The first n partial derivatives have the form:

$$\frac{\partial \bar{S}}{\partial Q_{it}} = P_{it} - V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) - \sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\partial D_t}{\partial Q_{it}}$$

The relation for the investment is written as follows:

$$\frac{\partial S}{\partial k} = -\sum_{t=1}^{T}\sum_{i=1}^{n} V_{it} Q_{it} \frac{\partial D_t}{\partial k} - \Psi'(k)$$

16

Thus, the first-order conditions lead to the following two relations:

$$P_{it} = V_{it} D_t \left(Q_1 \ldots Q_i \ldots Q_n, k\right) + \sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\bar{o} D_t}{\partial Q_{ti}} \quad \textbf{(1)} \text{ and}$$

$$\Psi'(k) = -\sum_{t=1}^{T} \sum_{i=1}^{n} V_{it} Q_{it} \frac{\partial D_t}{\partial k} \quad 17 \qquad \textbf{(2)}$$

Equation (1) establishes the principle of the pricing rule. The first term of this expression represents the „cost" accepted by the user i resulting from the loss of quality of service associated with congestion. The second term represents the value of the quality of service lost by all the users because of the last utilisation by the user i: this last term represents the external cost of congestion for which the latter is responsible.

Using classical pricing terminology, the optimal pricing is such that the price corresponds to the sum of the private marginal cost and the social marginal cost. The first term is already borne by the user. The toll to be applied should therefore correspond to:

$$\sum_{j=1etj\neq i}^{n} Q_{jt} V_{jt} \frac{\bar{o} D_t}{\partial Q_{ti}} \quad 18$$

If we now consider the second equation (2), which incorporates the investment dimension, it indicates that the capacity should be developed until the marginal investment is equivalent to all the congestion costs avoided.

Finally, and this is crucial, the optimum in terms of social efficiency will only be ensured if these two conditions are satisfied. The golden rule of marginal cost pricing consists of these two conditions, which theory suggests should be kept together. In common parlance, it could be said that this dual logic consists in establishing a price which, wherever there is saturation of the network, ensures either that demand is sufficiently well managed for the available capacity to be still sufficient or that the cost of the necessary investment in capacity can be covered. However, the quality of service production function may have different characteristics depending on the importance accorded to the intensity of the demand. The optimisation procedures will then be more or less determined by scarcity phenomena.

The congestion toll described above may take several forms, depending on the degree of divisibility of infrastructure use. The less the divisibility, as in the case of road transport, the more the intensity of use will affect the quality of service. But the greater the divisibility the less perceptible this effect and the more pricing should be oriented towards scarcity management systems, as in the case of car parks. The analysis should therefore be focused on the technical and economic relationship between the infrastructure and the uses of that infrastructure.

The rail system does not readily lend itself to this type of analysis. The existence of timetable slots seems to give credence to the idea that the use of the infrastructure is totally divisible and that there is therefore a strong user rivalry among the various consumers. Thus, it is theoretically possible to establish a market in user rights, for example, by auctioning slots, and thus reconcile supply and demand. It seems that this might be practicable on railway lines dedicated to similar kinds of traffic since, in this case, the slots auctioned would tend to be homogeneous. However, the only configuration that would seem to lend itself to this exercise is that of lines specialising in freight and open to several carriers, which could only apply to the lines, still to be organised, of a trans-European rail freight network. The first "freight corridors" to be established are not organised along these competitive lines.

In most cases, competition remains latent and overshadowed by slot allocation rules which predate the reforms. However, it is very real wherever network saturation is a problem, which brings us to the difficult question of capacity.

Rail capacity on part of the network can be very roughly and provisionally defined as the maximum possible number of movements that can be handled per hour. The reality, however, is much more complex.

On the one hand, capacity is determined by the characteristics of the infrastructure itself: the configuration of the lines and sets of tracks, the existence of community lines, switches, track intersections, the speed permitted by the design of the track, gradients, etc. Moreover, the capacity also depends on the utilisation of the infrastructure: type of trains (length, axle load, etc.), speed, number of stops, etc. The train schedule is then decisive and its organisation will have an impact on the network's effective capacity, which will be all the more considerable the more heterogeneous the slots concerned. The capacity will then depend on the train schedule as it has been drawn up. The scheduler and the rules of arbitration on which he relies will thus play an important role, a role which, in railway tradition, has not been much influenced by the considerations of the economist who, for his part, is keen to maximise the surplus that can be extracted from the infrastructure. To that end, he analyses the economic advantages which the consumers derive from using the infrastructure. For example, a decrease in the time taken by a train to travel between two stations may lead to a reduction in the technical capacity of the infrastructure by eliminating slots, but that reduction in capacity may be justified because it results in a net gain in surplus production. Conversely, the economist lacks mastery of the complex relations between the characteristics of the infrastructure, the possible operating programmes and the response of the final demand.

This difficulty involves both infrastructure pricing and project evaluation. Whether it is a question of calculating a developing marginal cost or the return on an increase in capacity, it should be kept clearly in mind that such an increase can very often be achieved in different ways: an additional junction, flyovers for avoiding track intersections, higher speed limits, passing track en route or in the station, improved block arrangement, better power supply, safety and traffic control installations, restoration of alternative routes, etc. As with the roads, the optimum infrastructure capacity is not a technical but an economic factor which, however, for rail forms part of a complex universe of technical solutions.

Here, we have one of the explanations for that common tendency of railway reforms which "separate" infrastructure management from transport production. The unified monopoly operator or separate operators not being omniscient, whatever the theoretical optimisation models may suppose, have not efforts been made to improve the efficiency of the system by giving the carrier the opportunity to make the best possible use of the signals sent him by the infrastructure manager and vice versa? Thus, it is for the carrier to interpret the infrastructure user price signal and for the infrastructure manager to interpret the demand signals he receives from the carrier. These "interpretations" could be based on the teachings of public economics.

However, to the above-mentioned difficulties we should add another associated with a familiar feature characteristic of most rail systems, namely, the presence of very considerable fixed costs and modest marginal costs over most of the network, i.e. wherever there are no saturation effects. The principles we have noted then lead to pricing which can result in a considerable deficit which it is the community's responsibility to meet. Whence the now acknowledged need to take this "budgetary constraint" into account.

3.3. Allowing for the budgetary constraint

The doctrine according to which the producer price of industries with diminishing costs should depend only on the marginal operating costs, and the rule according to which the authorities should cover all the fixed costs from taxes, first made an explicit appearance in the railway literature of the late nineteenth century. The French tradition of the economist engineers of the École des Ponts et Chaussées made a big contribution to these developments and their application. The debate became more animated following the appearance of Hotelling's article[45] in 1938. He concluded that the deficit resulting from the application of this global pricing principle should be financed by global taxes which, like the taxes on rent charges or inherited income, are supposed, in theory, not to affect the marginal behaviour of the economic agents.

Nevertheless, there may be a conflict between this theoretical viewpoint and another scarcity phenomenon, the public finances. Moreover, it assumes that resources are optimally allocated by the public operators, which is not necessarily the case when they believe themselves to be free of all financial constraints. It is to correct these dysfunctions that an entire theoretical school has devoted itself to justifying the addition of a budgetary constraint to marginal cost pricing. There are several ways of understanding this approach.

For some, this concern for a balanced budget is, in the words of Serge-Christophe Kolm, no more than the obsession of "*a short-sighted and narrow-minded accountant, who knows nothing of economics, trying mistakenly to transpose the criteria of the private to the public sector*[46]." By refusing to consider the main analytical conclusions of welfare economics, the supporters of the balanced budget deny themselves the possibility of satisfying, with the tools of economics, the demands of the public interest. Without rebutting this position directly, Mark Blaug has made it the target of some equally critical remarks which reflect the difficulties created by this pricing principle. Thus, according to Blaug, what the Anglo-Saxon authors call the French school[47] has trouble in taking the existence of deficits into account: "*The characteristic feature of the French contributions to the literature (on marginal cost pricing) is a total inability to take into account the problem of deficits in the diminishing cost industries which, indeed, hardly receives a mention*[48]."

For others, there are fewer disadvantages (i.e. less loss of global surplus) in distorting optimal pricing to control the deficit than in leaving the latter to drift while seeking to adhere strictly to the optimal pricing principle. Thus, in France and Europe, there has been a slow swing of the pendulum: the arguments in favour of marginal cost pricing no longer convince the authorities, who are more concerned about the financial situation of the public corporations, especially that of the railway companies which are steadily losing modal share, as well as about the unfortunate effects which the systematic covering of the deficits is having on the management of those enterprises.

Driven by the structural difficulties of the public finances, this trend is also based on the failure of a pricing system which requires transfers between taxpayers and users to ensure that the cost of the service to the community is explicitly weighed against the interests of those who use it. The gap between those who benefit from the system and those who finance it leaves a space within which the economic agents can conceal their preferences. The deficit subsidies may also hide and hence permit inefficient operation. Now, if there were no strong incentive to seek the minimum average cost, the willingness of the State to close any gap between the marginal cost and the average cost would result in enormous waste, which would be all the more enormous as it would probably be invisible and almost undetectable[49]. Thus the balanced budget constraint is aimed at an efficiency deficit which goes far beyond the traditional criticisms levelled against the champions of marginal cost pricing concerning the difficulties of evaluating the marginal cost and the technical, political and institutional barriers to implementation.

The following list of criticisms, without being exhaustive, will serve as an illustration: the pronounced indivisibilities of the infrastructure would lead to "sawtooth" pricing incompatible with long-term decision-making by the economic agents; the lack of rules for allocating certain cost elements would make competition impossible because of the existence of cross-subsidies; differential pricing on the network would bring into question an entire spatial standardization system; the practice would make it necessary for both users and authorities to gather, at great expense, information on the competitive structure of the market, on externalities, on the elasticity of the demand, etc. and, in short, the additional costs which the public would have to bear to implement these pricing systems would be out of all proportion to the advantages which they are supposed to bring.

As Vickrey points out, we are constantly on the horns of a dilemma from which it is difficult to escape completely. On the one hand, theoretically, the application of marginal cost pricing ensures that infrastructure utilisation is optimised but, considering the financial scarcity constraint, we then deprive ourselves of information on the real value of a new project or, if the project has already been carried out, about whether it is still worth operating. On the other hand, with the balanced budget constraint, we can be sure that the project or its operation are worthwhile, but we do not know whether the infrastructure is being utilised to best advantage[50]. More generally, by employing crude regulatory mechanisms, by excluding a number of users and by eliminating certain activities, the application of such a rule might lead to a serious loss of social efficiency[51].

All the solutions proposed for introducing a budgetary constraint into the pricing system must face this radical criticism: the allocation of the fixed costs to a particular user or a particular use remains largely arbitrary. Thus, the calculations form the subject of endless discussions about the relevance of a particular distribution scheme. It might seem that subtle cost accounting could reduce the proportion of non-apportionable costs and thus eliminate the problem. However, these methods can always be debated and offer no solution for a usually still significant residuum.

Now, from the standpoint of the optimal allocation of resources, the pricing system should not be mainly concerned with distributing the costs but, more fundamentally, should favour the achievement of a surplus. The objective is much more ambitious. Leaving aside the utilisation costs, for which the allocation procedures do not, in theory, pose any particular problems, it is clearly the availability of the infrastructure that must be reflected in the scales and not its effective consumption. Then, the efficient mode of contribution is not to seek to allocate costs but to find a means of extracting the surplus needed to finance the infrastructure while ensuring that the surplus is achieved. From the moment that this surplus exists (it is the role of public economic analysis to locate it), there exists a pricing system to bring it to light.

It is the segmentation of the demand which, in this respect, makes a vital contribution to economically efficient pricing. The theory of surpluses leads to precedence being given to pricing systems based on this latter principle. In practice, it may lead to the adoption of the monopolists' rule to the effect that the best pricing consists in imposing the charges that the traffic will bear.

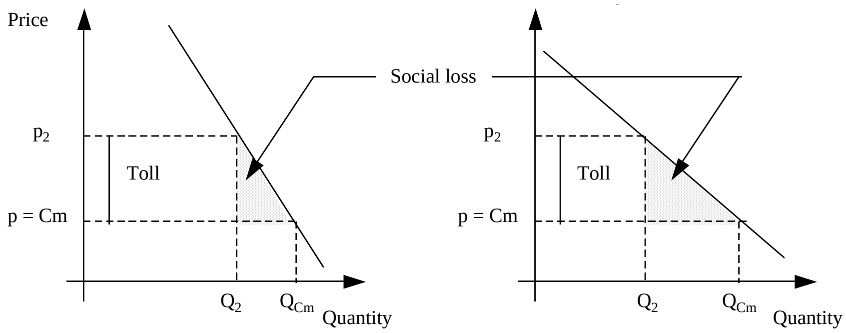
In fact, the question of the social loss involved in pricing that deviates from the marginal cost, like the question of the deficit and the limitation of its perverse effects, makes sense only if price unity is assumed. As soon as a discriminatory approach is adopted, the difficulty may disappear. Price discrimination, which takes into account the response to prices of each segment of demand, then makes it possible to increase the global surplus since the number of users will increase, while ensuring better coverage of the costs for the producer. The introduction of socially efficient price discrimination turns pricing completely upside down. It should be distinguished from the pricing principle, often encountered in the literature, that everyone pays his share, which makes even less sense inasmuch as the allocation of some costs is arbitrary. It forms part of another approach which authorises any use of a good calculated to generate a positive net surplus. Under this condition, not only will any discriminatory pricing system be neutral from the standpoint of the optimal allocation of resources, but it will be totally justified from the standpoint of the community since it will enable a surplus to be generated. Thus, the first principle considered to characterise optimal pricing gives way to another principle.

A pricing system is deemed to be economically satisfactory if the operator procures for each user a share in the absolute utility of the service sufficient to constitute an effective incentive for him to use the infrastructure. The contribution to this service is then said to be fair as long as, for each consumer and for each use, it does not exceed the net value of the utility he derives from it which, it should be recalled, must be positive. Although often disputed[52], this approach, based on tapping into a surplus, has a definite advantage in relation to the problem of financing which, as Jules Dupuit suggested, consists in "demanding as the price of the service provided not what it costs the provider but an amount commensurable with the importance attached to it by the one for whom it is provided[53].

This basic principle remains very theoretical and there are major difficulties to be overcome before it can be applied. However, the tools do exist. In 1956, Marcel Boiteux[54] proposed a solution which marked an epoch in the history of economic thought. The literature has even associated the name of the author with that solution so that it is customary to speak of Ramsey[55]-Boiteux pricing. Anglo-Saxon authors also refer to this principle as the "Inverse Elasticity Rule". Tradition has it that this seminal article offers a general solution of the problem of the production and the Pareto-optimal pricing of a public monopoly obliged to balance its budget within the context of a competitive economy.

Here, then, the decisive concept of elasticity of demand makes its appearance. Thus, if the enterprise is considered to have several markets at its disposal, observance of the budgetary rule which requires deviation from the marginal cost will lead to the prices for each market being so determined as to make the mark-up between the consumer price and the marginal cost proportional to the inverse of the price elasticity. In practice, this comes down to saddling the goods or services for which the demand is relatively inelastic with a larger share of the deficit. This pricing practice harks back intuitively to a mechanism well known in the field of taxation: the loss for the community resulting from the imposition of a tax (that is, in this case, the reduced consumption of a good consequent upon an increase in price) will be the greater the more elastic the demand.

Figure 3. **Social loss and elasticity of demand**



Assuming a monopolistic market on which the demand is very elastic, the slightest variation in price on this market will be reflected in substantial changes in the pattern of consumption. Thus, the so-called Boiteux rule consists in taking advantage of different relative market situations. The steeper the slope of the demand curve on a market, the lower the elasticity and the more limited the social loss resulting from a deviation from the marginal cost. Conversely, on a highly elastic market, the slope of the curve will be flatter. In this case, a deviation of the price from the marginal cost will be reflected in a heavy social loss.

Thus, when by necessity an enterprise has to cover the whole of its costs and hence, in the case in question, when it is obliged to deviate from the marginal cost, this theoretical demonstration makes it possible to justify placing the strain on the consumption which is the most inelastic. When the demand is relatively inelastic, the deviation of the Ramsey-Boiteux price from the marginal cost will be small and hence the deviation at the optimum will be minimised. The mark-ups will be greatest on the least sensitive demand. Thus, this method of pricing seeks not to distort the price signal sent to the most sensitive users in order that they may not significantly modify their pattern of consumption and to levy the charges on the less sensitive users who will not reduce their consumption more than slightly relative to the social optimum, even if the prices are raised.

Thus, returning to the allocation of non-apportionable fixed costs, where a monopoly can rely on several products it should parcel out its fixed costs according to the sensitivity of the demand. Prices are raised sharply where the demand is not sensitive and reduced where it is. The quantity consumed on each market remains as close as possible to the consumption which would have been observed in the first-best case. The optimum obtained maximises the social surplus subject to the constraint of a balanced company budget or, if this constraint seems too harsh or inaccessible, by assigning in the Ramsey-Boiteux optimisation programme a scarcity coefficient which overestimates the collective utility of the public contribution.

As Boiteux has himself been pointing out since 195656, to the difficulties of application of this pricing method there must be added the unrealistic assumptions of the model. At the same time, the author questions whether the practical application of the results obtained is of any real interest. In fact, the application of this rule poses a number of problems, in particular, by requiring a knowledge of demand elasticities. More generally, the necessary hypothesis of an omniscient economic supremo is obviously unrealistic. The fact that it is a question of pricing the use of the rail infrastructure and not the transport service itself is a further complication. The carrier is then the source of the demand, in the sense that he needs slots and he himself responds to a combined demand: that of the users or shippers and that of the transport organising authorities who „purchase" qualities of service from him. Thus, the effects of infrastructure pricing on the final demand are linked with the pricing principles which the carrier himself applies. For the system to be coherent and economically efficient, it is very likely that the two operators will have to price in accordance with similar principles, but this linkage has still to be subjected to a complete analysis.

This brings us to an essential conclusion concerning the pricing system which Maurice Allais considers to be one of the key elements distinguishing his theory of the markets economy from the standard model57. Within this more general theoretical framework, which seeks to maximise the collective surplus, the deficit constraint can be removed provided there is no objection to questioning price unity. That is an intuition already clearly expressed by Jules Dupuit58 in his time. If the surplus generated by an infrastructure is greater than the cost of putting it in place, then there necessarily exists a pricing system capable of tapping into this surplus to obtain the sums needed to finance the project while maintaining maximum social utility. The achievement of the surplus becomes the challenge of the pricing procedures. It can justify the transition from a logic of differentiation based on costs alone to a logic of discrimination based on segmentation of the demand. It is not a question of differentiating prices solely on a user pays basis, which makes little sense since, once again, a portion of the costs cannot be distributed in accordance with this principle. The discrimination of the demand should be based on the principle of he pays who can, especially when that is the only way of obtaining a return on projects with a positive global net utility. Otherwise, if the project is not financed out of the general budget, a type of solution now considered best avoided, it will never materialise.

We then enter a pricing universe that is more complex but still regulated by marginal cost in which it is less a matter of minimising the effect of a deficit linked to optimal pricing as of seeking to maximise the surplus by differentiating the pricing, segmenting the customer base and, finally, obliging all the economic actors to disclose their preferences. These principles now need to be compared with the actual results of implementing the most significant recent reforms.

4. APPLICATION OF PRINCIPLES AND NETWORK CONSTRAINTS

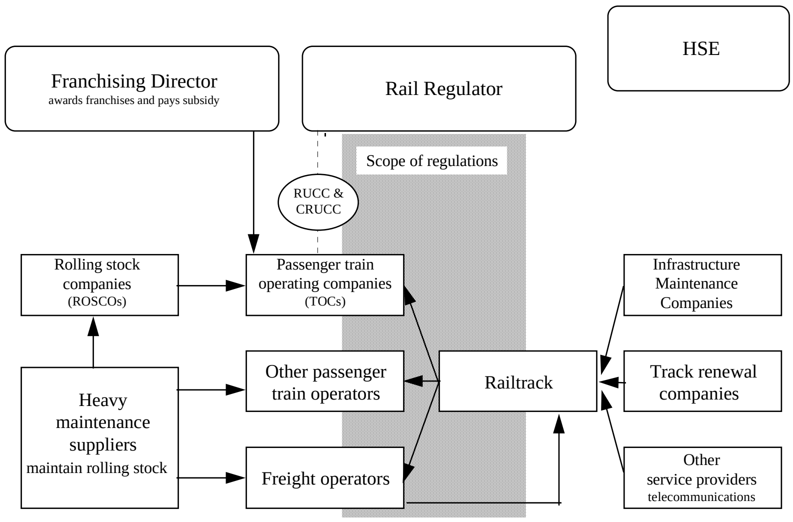
In this final part of our paper, we shall compare the few theoretical principles that can be derived from economic analysis with the practices of networks which, as a result of having been reformed, should have solved the infrastructure pricing problem. Accordingly, we shall examine, in turn, the British, German and French cases.

4.1. The British experience: making a surplus59

The restructuring of British Railways began in the early 80s. This long process, intended to improve the efficiency of the sector, led the authorities gradually to modify the enterprise's internal organisation. On 1 April 1994, the reforms took a further step forward with the entry into force of the Railways Act which prepared the ground for a phased privatisation of the sector by programming, in particular, for the separation of operations from infrastructure. The enterprise's transport services have thus been separated into smaller entities which, from the outset, the authorities have made clear they intend to privatise. As for the infrastructure, it has been assigned to a new, private-law company held by the State, Railtrack. This company, still a monopoly, retains operational control over the traffic, allocates capacity and, above all, is responsible for pricing use and determining the fees to be paid by the various operators to ensure that the costs are covered60. The company, which initially was left in public hands and even benefited from investment subsidies, was privatised in 1996. BR's passenger transport business has been split up into 25 separate entities which have been placed under private sector control by introducing a franchising system. The freight business has all been sold off to the private sector and opened up to competition. Thus, Railtrack buys services from and sells them to a range of operators.

At the same time as splitting up BR, the Railways Act established a powerful regulatory system based on three bodies. The first, the Office of the Rail Regulator (ORR), is mainly responsible for supervising infrastructure access and pricing. It establishes the rules of competition and oversees their application, especially in the interests of the customer. The task of the Office of Passenger Rail Franchising (OPRAF) is to grant franchises and supervise the correct application of the terms and conditions by each franchise holder, in particular as regards the consistency of the services actually provided. Finally, the Health and Safety Executive (HSE) ensures that the safety regulations are observed. It issues rules governing the design, construction and operation of rolling stock, infrastructure and equipment.

Figure 4. **The British reforms**



Source: ORR Report, 1996-97.

Each operator signs a track access agreement with Railtrack. Two sectors should be distinguished. Initially (7-15 years), the franchised passenger line operators have been granted the access rights necessary to provide the services stipulated in their specifications. On certain routes they enjoy protection which will eventually be withdrawn. The existing freight carriers have been granted initial rights to enable them to satisfy their present customers. Apart from the time slots allocated in connection with these rights, there are others which are open to competition.

The splitting up of the network immediately posed the problem of the allocation of costs among the various activities. It was decided that the costs should be allocated to the various sectors in such a way that each sector bore the costs of the fixed assets and personnel of which he was deemed to be the principal user. The basis for this pricing is that it must be sufficient to ensure Railtrack a certain return on its assets.

The rail operators pay Railtrack infrastructure user fees intended to cover the network utilisation and signalling costs and the cost of supplying power, where appropriate. Thus, the overall aim is to pursue a balanced-budget pricing policy which also takes into account the ability to pay of the applicants for time slots. Accordingly, it is not possible to speak of Ramsey-Boiteux pricing because of the special terms granted to the franchise operators.

In fact, the pricing system differs depending on whether one considers the operators who provide services under franchise (subsidised services) or the operators who purchase time slots on the network. The companies which operate passenger lines under licence are in a special situation since in this case the structure and the level of the access charges applicable are directly controlled by the regulatory body. The access charges applicable to the passenger lines operated under licence[61] are based on a cost allocation study that uses the concepts of avoidable[62] and additional[63] costs.

In the event of a time slot purchase, Railtrack is free to negotiate its prices, although the contract must be approved by the Regulator. The fees are negotiated but subject to approval based on the principles established by the Office of the Regulator.

The general principle requires that the fee structure should not deviate too far from the value of network access for the users and that it should enable Railtrack to recover all the costs actually incurred in connection with the transport of goods, to which there should be added a possible contribution to cover the joint costs shared with the passenger services. Thus, the pricing rule should be such that the minimum price is not less than the avoidable costs occasioned by the service concerned. The price should be less than or equal to the costs which Railtrack would incur if the operator were alone on that portion of the network and then had to assume all the costs. The price should not, when the various cost factors are taken into account, differ appreciably from one user to another.

In addition to the transparency of the charges being more difficult to achieve for freight than for passenger transport and it being difficult to determine precisely the share of the cost directly apportionable to a particular service, the allocation of the joint costs is a real headache.

The reforms have been the target of two main criticisms. Some consider that Railtrack, by setting low prices for freight transport, for which it is competing, may seek to shift the fixed costs onto the passengers and thus improperly obtain public funding to its advantage. Others, on the other hand, consider that the draconian safety regulations, though reasonable in the case of passenger transport, are less justified in the case of freight transport and, accordingly, that there has been a transfer from the freight to the passenger sector.

However, in 1995, before Railtrack was privatised, after examining the access charges applied to the franchised passenger services, the ORR concluded that they were too high and more than the operator really needed to fulfil his infrastructure renewal obligations. The Regulator imposed corresponding price reductions, thus transferring the productivity gains to the licensed operators. The charges are to be reviewed in the year 2000. Another object of criticism is the charge structure, the fixed portion being considered too high (91 per cent). This pricing is not conducive to the rational management of resources. In fact, this approach precludes the introduction into the access charge calculations of differentials in terms of rush hours or the economic value of the slots. The costs considered here do not incorporate such externalities as noise or air pollution.

When the pricing rules were being drawn up, there was a keen debate between the advocates of a commercial strategy and those who favoured a more managed approach. The arguments are important inasmuch as the same debate is now being conducted at the European level. After holding numerous consultations, the ORR concluded that it was better to place the method of calculating the infrastructure user fees on a commercial negotiation basis so as to give the operator a chance to attract all the economic agents capable of paying at least the directly apportionable costs. Thus, the main aim is to give the infrastructure manager the means to induce the maximum possible number of operators to use the network. Clearly, then, the objective is to achieve and engineer a surplus. This approach has been much criticised. The owner of the infrastructure can engage in cross-subsidisation and favour one party or another without necessarily ensuring the complete opening up of the network

4.2 The German experience: covering the costs

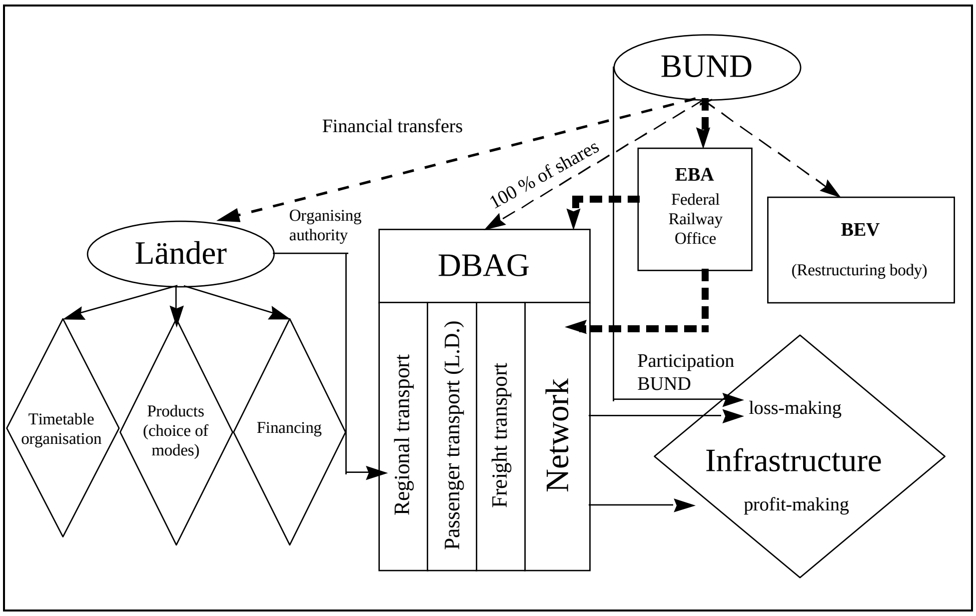
As in other countries, the amount of federal subsidies granted to the rail sector in order to balance its budget and guarantee its borrowings had become considerable while, at the same time, rail's share of total traffic was being continually eroded. Reunification only worsened the crisis and, in 1993, the enterprise (made up of the DB and the RB) had debts of more than DM 67 billion while the Government was anxious to support the development of the sector. The reform of German railways was speeded up, profoundly transforming the rail transport situation across the Rhine. It opened the way for ever keener competition on the railways, on the one hand in the short term, by placing the historical rail operator in the position of a service provider in its negotiations with each Land and, on the other, in the long term, by opening up the network to third parties. The reforms entered into effect, following an amendment to the Constitution, on 1 January 1994.

Thus, the authorities gradually moved towards a vast controlled and concerted structural reform of the railway company. The Central Government (*Bund*) set up a private joint-stock company, the DB AG, in which at present it holds all the shares but which is supposed to disappear in 1999. The rail system is organised in four independent sectors of activity: regional passenger transport, long-distance passenger transport, freight transport and infrastructure management. These sectors are eventually to be privatised. However, the Federal Government will remain the principal shareholder in the *Fahrweg* (infrastructure manager) in order to retain control over investment policy.

As in the United Kingdom, the reforms are based on new institutions. On the one hand, the Federal Office of Railways - the *Eisenbahnbundesamt* (EBA) - has been set up to ensure the necessary co-ordination and take care of the general missions of the railways. It authorises operations, certifies equipment and organises work on the infrastructure. It plans the work on the federal rail network, ensures that tendering procedures comply with the law, grants licences, applies investment financing agreements, commences prosecutions and settles disputes. On the other hand, the *Bundeseisenbahnvermögen* (BEV), another federal body, is responsible for clearing the debts of the former company and for administering staff and pension costs as well as financial and property charges. Central government has taken over the costs of the staff of the former DB with civil servant status, who can now be made available to the DB AG for employment under ordinary market conditions.

At the same time, to meet the increasing costs of regional rail transport, the central authorities have transferred to the *Länder*, with their agreement, the responsibility for organising and financing regional transport operations. Thus, the latter will henceforth find themselves in the position of organising authority. They will receive financial transfers from the Bund to enable them to perform this task64.

Figure 5. **The German reforms**



Infrastructure pricing is an important component of these reforms. The Act authorises access for third parties and thus transposes into domestic law the principles of Directive 91/440. The operating subsidiaries of the DB AG pay the infrastructure manager for the use of the infrastructure. These charges are published in a catalogue which tells the operators exactly what price they will have to pay for the whole of a journey, depending on their requirements. In July 1994, DB Infrastructure published an initial price schedule applicable to all users of the network. The system was chopped up into 4 000 sections with well defined characteristics on the following principles.

The prices are identical as between the DB AG and third parties and must be identical for requests with similar requirements. The differentiation to be found in the catalogues is based on objective criteria. The prices depend on the category of line, the damage potential of the equipment used, which largely depends on the type of use, the required regularity, the volume of purchases and the length of time for which the slot is used. The basic tables are compiled on the basis of four criteria, namely, the quality of the track (essentially the permissible speed), the traffic potential (principal characteristics of the rail links) according to the type of service requested, the wear and tear (based on a variable cost analysis) and the planning quality. This last item relates to the quality of service demanded, the reliability (punctuality) indicator determining the room for manoeuvre left to the infrastructure manager. This indicator is expressed in the form of a percentage representing the margin which the DB is allowed relative to the theoretical journey time. Using these criteria, it is possible to construct the reference table shown in Table 1.

The notion of quality introduces an element of economic demand management. Here, the aim of differentiation is to define a scale of operator's requirements. In choosing a quality of service, the operators disclose their preferences. However, there are strict limits to this mechanism since the percentage is fixed for each category. It would be different if the operator could choose a level of reliability in each category.

Many weightings and modes of payment can be introduced on the basis of this table. If the stated maximum load is exceeded, the basic slot price is increased by 1 per cent for every additional 100 tons.

For trains running empty, the basic slot price for price classes P1 to P3 is reduced by 10 per cent and that for price classes P4 to P7 by 5 per cent. For engines running light, the basic slot price is reduced by 20 per cent.

Every regular slot ordered must be paid for irrespective of its utilisation. It is possible to reserve optional slots. A reservation fee of 20 per cent of the slot price is then collected. This fee is non-returnable.

Customers who order many slots are granted a price reduction which depends on the annual total of train-kilometres.

---

1. The pricing of the various transport sector modes should also take into account the environmental social costs and propose coherent pricing rules. This presentation does not expand on this point, which would require special development. We note that in this respect it is necessary to distinguish between two independent types of questions. On the one hand, rail pricing should be considered in relation to the social costs actually taken into account in the pricing of the principal competing mode, namely, road transport. This concern for coherence might even lead to the legitimising of intermodal balancing subsidies, theoretically justified by a second-order optimum. On the other hand, and this is a totally different problem, it is necessary to introduce differential pricing within the rail mode in order to take into account the advantages and disadvantages of the different technologies used by the operators and gradually encourage the use of those that are less polluting.

2. Brooks, M. and Button, K. propose a detailed typology of these various experiences in "Separating transport track from operations: a typology of international experiences", *International Journal of Transport Economics*, Vol. XXII, No. 3, October 1995.

3. This is the case in the United States for certain passenger services for which Amtrak must negotiate an infrastructure access charge with the integrated private operators.

4. The calculation methods used in these studies can often be reduced to very simple procedures, especially as the quality of data needed for more sophisticated calculations is very difficult to obtain. Nevertheless, the statistical analysis can be made more precise by greater refinement and by establishing precise relations between particular types of costs, networks and users.

5. Avoidable fixed costs are then defined as those which would disappear if the firm stopped producing one of its products.

6. Winston, in a survey of road pricing, explains that the proposal to consider optimal pricing and optimal investment in parallel harks back to the work of Herbert Mohring (1962).

This school includes such authors as T.E. Keeler, K.A. Small, M. Kraus, S. Glaister and S.A. Morrison.

Morrison, S.A. (1986), "A Survey of Road Pricing", *Transportation Research*, Vol. 20A, No. 2, March, pp. 85-97.

Mohring, H. and Harwitz, M.I. (1962), *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston, IL.

Small, Kenneth A., Clifford Winston and Carol A. Evans (1989), *Road Work: A new Highway Pricing and Investment Policy*, Washington, D.C., The Brookings Institution.

Hau, Timothy D. (1992), *Economic Fundamentals of Road Pricing: A Diagrammatic Analysis*, Transport Division, Infrastructure and Urban Development Department, The World Bank, December.

7. Allais, Maurice (1948), *Revue d'Économie Politique*, *op. cit*., p. 230.

8. Boiteux, Marcel (1994), "Le développement de l'approche économique du service public", *L'Europe à l'épreuve de l'intérêt général*, Collection ISUPE, Édition ASPE EUROPE, Paris, p. 46.

9. Kolm, Serge-Christophe (1968), "La théorie économique générale de l'encombrement", *Futuribles*, SEDEIS.

10. Although traffic management is generally based on rules of priority, which distinguishes the rail problem from that of the roads and creates a further difficulty for the theoretical approach to the pricing of rail infrastructure.

11. Here we have used the demonstration proposed by Steven A. Morrison, who bases himself on the work of authors such as Mohring, Harwitz and Vickrey.

12. Hotelling, H. (1938), "The General Welfare in relation to problems of Taxation and of Railways and Utility Rates", *Econometrica*, 6(3), July, pp. 242-269.

13. Kolm, Serge-Christophe (1971), *La valeur publique, l'État et le système des prix,* CNRS, Dunod, Paris, p. 7.

14. In the literature, the French school is associated with the rejection of pricing based solely on balancing the accounts. This takes little account of the analyses of Jules Dupuit in the last century or of the more recent work of Maurice Allais and Marcel Boiteux.

15. Blaug, Mark (1986), "La pensée économique, origine et développement", *Economica*, 4th Edition, Paris, p. 718.

16. Allais, Maurice (1948), "Le problème de la coordination des transports et la théorie économique", *Revue d'Économie politique*, 58, pp. 212-271.

17. Vickrey, W. (1948), "Some Objections to Marginal Cost Pricing", in

accountability of the managers. In no circumstances should this second factor be neglected." *Note Prévision*, Commissariat Général du Plan, p. 151.

18. Bernard, A. (1983), "Coût marginal ou coût moyen ?", in Quinet, E. (editor), *Les transports et la puissance publique*, Presses de l'École Nationale des Ponts et Chaussées, report on transport economics seminar organised in 1981-82 by the École Nationale des Ponts et Chaussées, p. 36.

19. Very many commentaries on this type of pricing point out that fairness implies that the consumers should bear the costs of producing the goods they consume and that all the consumers should pay the same unit price for the same good. This is a frequently recurring complaint: "those systems which differentiate between deficit tolls according to the characteristics of the demand are generally considered unreasonable and unfair". Oort, C.J., *op. cit*., p. 62.

Clearly, the theoretical considerations advanced here shatter this principle. The tolls applied may vary for products that are identical both from the technical standpoint and by reason of their cost.

20. Dupuit, Jules (1849), "De l'influence des péages sur l'utilité des voies de communication", *Annales des ponts et chaussées*, p. 248.

21. Boiteux, Marcel, "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire", *Econometrica*, No. 24, 1956, pp. 22-40.

22. Ramsey, F. (1927), "A contribution to the theory of taxation", *Economic Journal*, Vol. 37, No. 1.

24. With that "s" Maurice Allais distances himself from the standard model. For further details, see: Allais, Maurice (1971), "Les théories de l'équilibre économique général et de l'efficacité maximale. Impasses récentes et nouvelles perspectives", *Revue d'Économie Politique*, No. 3, May-June, pp. 331-409.

25. Dupuit, Jules (1849), "De l'influence des péages sur l'utilité des voies de communication", *Annales des ponts et chaussées*, pp. 170-248.

26. We shall not describe the reforms themselves as they have already been extensively analysed. See, for example:

- Montagu (1995), "L'accès à l'infrastructure ferroviaire, l'expérience britannique", *Rail International*, 26, January, pp. 5-15. Paper read at the AICCF/CCFE Conference, Berlin, 6-7 October 1994: The Railways and European Transport Policy.
- (1994), *Britain's Railways: a new Era*, The Department of Transport, London, March.
- Preston, John (1996), "The economics of British Rail Privatisation: an assessment", *Transport Review*, Vol. 16, No. 1, pp. 1-21.
- British Ministry of Transport (1996), *Transport Report*.

  Available on the Web at:

  <URL http://www.open.gov.uk/dot/ann_rpt/chap_1.htm>.
- Nash, Christopher (1996), "The British Experience", *The Separation of Operations from Infrastructure in the Provision of Railway Services*, Round Table 103, ECMT, pp. 59-102.
- Bradshaw, B. and L. Mason (1994), *Rail Privatisation: Facts, Issues and Opportunities,* Oxford Economic Research Associates Ltd., Oxford.

27. Railtrack derives its income from user fees paid by the operators (supply of electricity, etc.), rents paid for the use of stations and depots and rents from its commercial assets. To these should be added the access charges which are determined by negotiation (see below). The procedures have been progressively refined. At the beginning, no rules for calculating the charges were laid down. The first charges were fixed at a level that would cover the total costs and ensure a return on the invested capital of the order of 8 per cent.

28. The charge includes a fixed annual fee comprising the allocated fixed costs (joint costs) and the additional fixed costs (specific to each company). The fixed charges, which correspond to about three-quarters of the infrastructure costs, form the subject of negotiations between the operators and Railtrack. The variable charges contain infrastructure user fees calculated in terms of train-miles which are different for each category of rolling stock (10 per cent of total costs). The costs incurred at regional and national levels are shared out among all the franchise holders in proportion to their receipts from fares. The costs incurred at local level, or on a single line, must be distributed among the users in proportion to the number of vehicle-kilometres travelled.

29. Avoidable costs: rule for the allocation of the fixed costs of the whole of the services provided by an operator, equal to the amount saved in the event of his services being eliminated.

30. Additional costs: increase in infrastructure costs imposed by its services, taking into account the configuration of the other services.

31. The corresponding subsidies are financed from revenue generated by the petroleum tax *(Mineralölsteuer).* Note that article 4 of the Railways Restructuring Act states that, from 1996, the DB AG will no longer receive any direct funds from the Federal Government for managing regional passenger services. The subsidies are allocated to the *Länder,* which use them in accordance with their own regional transport policy. However, the *Länder* must use these transfers for public transport purposes.

Table 1. **Quality of service reference table**

Price class	P1 High-speed traffic	P2 Express passenger traffic (main lines)	P3 Regional express passenger traffic	P4 Average-speed main-line passenger traffic	P5 Short-haul regional passenger traffic	P6 Local passenger traffic	P7 S-Bahn (urban rail transport)
Max. load (t)	1 000	750	550	750	400	400	450
Permissible speed on at least one section	200 km/h or more	Up to 200 km/h	Up to 160 km/h	Up to 140 km/h			
Planning quality	105 %	108 %	110 %	120 %	120 %	120 %	108 %
Category of train	Intercity express (ICE)	EuroCity and InterCity	InterRegio, main-line express trains	Express night trains, accompanied-car trains, fair trains	Regional express train, through train	Regional train, CityBahn, slow train	S-Bahn train

Table 2 **Percentage reduction in terms of train-km per year**

Percent reduction	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Main-line traffic train-km (millions)	14	28	42	56	70	84	98	112	126	140	154	168	182	196	210	224	238	252	266	280
Short-haul traffic train-km (millions)	0.3	3	25	63	134	205	250	293	333	370	407	444	481	518	555	592	629	666	703	740

Customers who order slots for several years and sign a contract are granted a further price reduction in addition to that mentioned above. For firm orders extending over 2, 3, 4 and 5 years, the corresponding reductions are 2, 3, 4 and 6 per cent.

The German pricing policy makes the financing of rail activities truly transparent, even though the determination of the costs in terms of train-km is far from receiving unanimous approval and constitutes an obstacle to the entry of new operators. The relatively high prices and the choice of pricing applied to the train rather than the wagon are dissuading new operators from moving in.

These price scales have introduced a certain flexibility, but it is still insufficient and, in a way, is institutionalising the status quo by discouraging the adoption of new techniques and limiting the freedom of manoeuvre of possible new entrants. Separation has not progressed very far because DB AG is still both service provider and network operator. The transition is a gentle one. The undertaking seems to have been genuinely successful since already more than sixty transport operators have moved onto the DB AG's rail network and their number is steadily increasing

While proposing rather high marginal network entry costs, this pricing system reduces, on the one hand, the uncertainty for future operators by encouraging longer-term commitments and a search for financial and technical partners and, on the other hand, short-term practices which could have pernicious effects on the continuity and quality of the rail service.

Finally, the choice of a high level of global pricing guarantees the infrastructure manager a development capability, which is one of the strong political choices of these reforms, together with the choice of a user tariff likely to lead to the optimal utilisation of the network.

4.3. The French example: a transitional phase[65]

Introduced against a background of social strife, in particular a big strike in 1995, the French reforms consisted in establishing *Réseau Ferré de France* (RFF), a new public company which, as balance-sheet liabilities, received three-quarters of the debt of the SNCF and, on the asset side, the national network infrastructure, with the exception of the stations and installations needed by the historical operator. The latter was entrusted with the management and maintenance of the infrastructure on behalf of the RFF, which pays the bill for this service (16.8 billion francs for 1997, the first year of implementation). The SNCF pays the RFF for the use of the infrastructure. For the first two years, a limit was placed on this fee (slightly under 6 billion in 1997) by the law and the decree establishing the new system.

The first characteristic of the system relates to the fact that, relieved of most of its debt and infrastructure financing, the SNCF is in a position to balance its accounts, which it is expected to do in 1999. Secondly, the new infrastructure company, which at present can only count on earning 6 billion francs, must cover, in addition to nearly 17 billion in network maintenance and management costs and unavoidable investment costs amounting to about 13.6 billion, charges of around 9 billion on the debt inherited from the reforms. Obviously, most of the difference between expenditure and income is covered by the Government, in the form of either a capital grant or subsidies, the difference being made up by borrowing.

This, then, is a very special situation which can be interpreted in two different ways. Either the RFF may be regarded as a body whose principal function is to take over the debt and cover the deficit (net of subsidies) of the infrastructure account by borrowing. Naturally, in this case, investment would be the adjustment variable and would inevitably face historical decline. Or the present situation may be regarded as a transitional phase for putting new structures in place, after which pricing that offers greater incentives and ensures better coverage of the costs will be gradually introduced.

Being capped in 1997 and 1998, the present fee system is obviously far removed from the principle of covering the costs. It corresponds to about one-fifth (in terms of the total amount) of the German system. There is little connection between the six billion constraint imposed and the reality of the actual costs, particularly as more than half of this sum comes from the regional organising authorities (mainly "Parisian" passenger transport) and less than half from the SNCF.

There is little point in studying this provisional system, precisely because it is capped. However, it should be noted that the idea was to create incentives, especially where the demand for slots is high relative to capacity, i.e. on the urban and suburban lines (the part of the network designated R_0) and to a lesser extent on the busy high-speed lines (R_1). On the other hand, on the low-density, high-speed network and on the main-line network (R_2) the fees are very low, while on the rest of the network (R_3) they are symbolic.

The fee system distinguishes between a monthly access charge, AC, per kilometre of lines for which access is requested, a reservation charge, RC, per kilometre and per slot reserved and a traffic charge, TC, per train-kilometre. There are different reservation charges for peak periods, normal periods and slack periods. The corresponding charges for 1997 are shown in Table 3.

Clearly, this system is much less detailed and sophisticated than that introduced on the German network and thus raises the question of whether it is sufficiently refined to enable the relevant marginal costs and homogeneous demand segments to be distinguished in the event of the future system being steered towards a more determinedly economic form of pricing.

Table 3. **French provisional pricing**

(1997 unit prices in francs)

Sub-network	**R_0**	**R_1**	**R_2**	**R_3**
AC	11 000	11 000	250	0
RC (peak)	100	18	0.85	0
RC (normal)	44	6	0.85	0
RC (slack)	20	4	0.85	0
TC	0.30	0.30	0.30	0.30

Note: AC is expressed per month and per km of line, RC per slot-km, TC per train-km.

In the French case, clearly the main problem is how pricing will evolve after 1999. This question is overshadowed by the fact that, overall, the rail system is running at a loss. The first step then will be to choose between two strategic directions. One choice would be a low-toll system which would concentrate the public contribution on covering the deficits of the RFF and financing new investment. In this case a policy of long-term marginal cost pricing without budgetary constraint might be envisaged. A second choice would be a system that combined budgetary constraint with a Ramsey-Boiteux principle. In this case the SNCF would have to be subsidised for a fairly long time, but the subsidies could be correlated with the loss-making services thus financed, thereby allowing the authorities latitude to compare their cost and their utility.

On the basis of a study in progress, the RFF is to propose to the Government a user pricing system designed to encourage a better allocation of resources.

5. CONCLUSION

To reach a conclusion on such a subject would be to suppose that a definitive theoretical contribution, which was both coherent and pertinent and proposed measurable concepts, would make it possible to solve, down to a few details, this difficult rail infrastructure pricing problem. Only a patient approach that takes into account all the attempts to apply theoretical prescriptions will enable us to work towards a satisfactory solution.

It is no insult to the achievements of economic theory or railway economics to conclude with the following few lines which were written about a century and a half ago:

We merely wished to show that the way in which the tolls are fixed can greatly extend the utility of certain routes and that the guiding principle in assessing these charges should not be to set a price proportional to the weight or the distance nor to favour a particular industry or a particular class of passengers, but rather to impose on each passenger and on each good only a price that is lower than that which would prevent the passenger or good from using the route. Admittedly, the methodical classification of these passengers and goods does call for inventiveness and an intimate knowledge of the local circumstances, but a sound theory can do much to facilitate this work[66].

NOTES

32. See the following official texts:

- Law No. 97-135 of 13 February 1997, establishing the public corporation, *Réseau Ferré de France*, with a view to the revival of rail transport;
- Decree No. 97-446 of 5 May 1997 on national rail network user fees;
- and, finally, the Orders of 30 December 1997 on national rail network user fees, JO, 31 December 1997, pp. 19461-19463.

33. Dupuit, Jules (1873), "Poids et mesure", in Coquelin and Guillaumin (eds.), *Dictionnaire de l'économie politique*, 4th Edition, Vol. 2, Paris, pp. 339-344.

BIBLIOGRAPHY

Allais, Maurice (1945), *Économie pure et rendement social*, Edition Sirey, Paris, 72 pp.

Allais, Maurice (1948), "Le problème de la coordination des transports et la théorie économique", *Revue d'Économie politique*, 58, pp. 212-271.

Allais, Maurice *et al.* (1965), Options de la politique tarifaire dans les transports, Report prepared at the request of Commission of the European Economic Community, *Studies Coll., Transport Series, No. 1*, Brussels.

Allais, Maurice (1989), "L'économie des infrastructures de transport et les fondements du calcul économique", *Revue d'Économie Politique*, 99, No. 2, pp. 159-197.

Allais, Maurice (1989), *Théorie générale des surplus*, PUG, Presse Universitaire de Grenoble (2nd edition), 716 pp.

Banister, David, Kenneth Button *et al.* (1993), *Transport, the Environment and Sustainable Development,* Alexandrine Press, Oxford, 275 pp.

Baumstark, L. (1997), "Tarification de l'usage des infrastructures et théorie de l'allocation optimale des ressources, de la logique des coûts à la révélation des préférences", Doctoral thesis, Université Lumière Lyons 2.

Boiteux, Marcel (1956), "La vente au coût marginal", *Revue Française de l'Énergie*, December.

Boiteux, Marcel (1956), "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire", *Econometrica*, No. 24, pp. 22-40.

Bowers, Ph. (1995), "Commercialisation, accès local et international aux réseaux ferrés", *Public Transport International,* No. 1.

Brooks, M., K. Button (1995), "Separating transport track from operations: a typology of international experiences", *International Journal of Transport Economics*, Vol. XXII, No. 3, October.

Council Directive 91/440 of 29 July 1991 on the development of the Community's railways, OJ L 237 of 24/08/1991, 25 pp.

Council Directive 95/18/EC of 19 June 1995 on the licensing of railway undertakings (OJ L 143/70-74).

Council Directive 95/19/EC of 19 June 1995 on the allocation of railway infrastructure capacity and the charging of infrastructure fees (OJ L 143/75-78), 27 June 1995.

Regulation (EEC) No. 1191/69 of the Council of 28 June 1969 (OJ L 156) on action by Member States concerning the obligations inherent in the concept of a public service in transport by rail, road and inland waterway, last amended by Regulation No. 1893/91 of the Council, OJ L 169 of 29/6/1991.

Regulation (EEC) 1192/69 of the Council of 26 June 1969 on common rules for the normalisation of the accounts of railway undertakings, OJ L 156 of 28/06/1969.

Regulation (EEC) No. 1107/70 of the Council of 4 June 1970 on the granting of aids for transport by rail, road and inland waterway, OJ L 130 of 15/06/1970.

Green Paper (1995), *Towards fair and efficient pricing in transport,* COM(95)691 of 20/12/1995, 58 pp.

White Paper (1996), *A strategy for revitalising the Community's railways*, COM(96) 421 final, Brussels, 30 July 1996, 54 pp.

ECMT (1995), *Why Do We Need Railways?,* Proceedings of International Seminar, Paris, 19-20 January, 226 pp.

ECMT (1993), *Privatisation of Railways, Round Table 90*, Paris.

ECMT (1997), *The Separation of Operations from Infrastructure in the Provision of Railway Services, Round Table 103*, Paris.

Commissariat Général du Plan (1995), *Transport, le prix d'une stratégie* (Workshop chaired by A. Bonnafous), Vols. 1 and 2, La Documentation Française, Paris.

Commissariat Général du Plan (1996), *Transformation des opérateurs publics de réseaux en Allemagne -- Les cas du contrôle aérien, des transports ferroviaires, de la poste et des télécommunications.* Study made by Cabinet ESL & Network, Vols. 1 and 2, March.

Commissariat Général du Plan (1993), *Transports : pour une cohérence stratégique*, Workshop on the strategic directions of transport policy and their medium-term implications, Commissariat Général du Plan, September.

Commissariat Général du Plan (1994), *Transports : pour un meilleur choix des investissements* (Group chaired by Marcel Boiteux), La Documentation Française, November, 132 pp.

Commissariat Général du Plan (1995), *Transports : le prix d'une stratégie* (Workshop chaired by Alain Bonnafous), Vols. 1 and 2, La Documentation Française, pp. 213 and 182.

DB AG (1996), "Der Trassenpreiskatalog der DB AG: ein Vorbild für Europa?" Lecture given by Mrs Haase at the 29th Scientific Transport Research Seminar, "Competition Policy in Deregulated Transport Markets", organised by the University of Freiburg (Germany), Hinterzarten, October.

DB AG, *Trassenpreise - Güterverkehr*, Frankfurt am Main, 1995.

DB AG, *Trassenpreise - Personenverkehr*, Frankfurt am Main, 1995.

Dodgson, J. (1994), "Railway privatisation", in: M. Bishop, J. Kay and C. Mayer (eds.), *Privatisation and Economic Performance*, Oxford University Press, Oxford.

Drèze, H. (1994), "Marginal-Cost Pricing", Introduction to Part III in: Arnott, Richard, Kenneth Arrow, Anthony B. Atkinson and Jacques Dreze (eds.), *Public Economics* (Selected papers by William Vickrey), Cambridge University Press, pp. 190-195.

Ewers, H.-J. (1993), Privatisierung und Deregulierung bei den Eisenbahnen: das Beispiel der DB und der DR, Institut für Verkehrswissenschaft, Universität Münster, 15-17 September, 11 pp.

Gillen, David W. and Tae Hoon Oum (1991), *Transportation infrastructure: Pricing, Investment and Cost Recovery,* Royal Commission on National Passenger Transportation, Technical Seminar, Ottawa, Canada, September 26 and 27, 181 pp.

Guihéry, L. (1996), "Fédéralisme allemand et réforme des chemins de fer en Allemagne : le transport ferroviaire régional de voyageurs devient une nouvelle compétence des Länder", *Documents-revue des questions allemandes*, No. 1, Paris, p. 79.

Guihéry, L. (1997), "Réforme des chemins de fer et tarification d'usage des infrastructures ferroviaires en Allemagne, une organisation contrôlée de la concurrence sur le réseau ferré allemand", *Transport,* No. 382, March-April, pp. 112-121.

Hau, Timothy D. (1993), "Economic Fundamentals of Road Pricing: A Diagrammatic Analysis", Working Paper, WPS 1070, Transport Division, Infrastructure and Urban Development Department, The World Bank, p. 87.

Horton, Sir Robert (1997), "Re-engineering Britain's railways: risks and rewards", *European Railway Review*, June, pp. 89-95.

Hotelling, H. (1938), "The general welfare in relation to problems of taxation and of railways and utility rates", *Econometrica*, 6(3), July, pp. 242-69.

Hutter, R. (1950), "La théorie économique et la gestion commerciale du chemin de fer", *Revue Générale des Chemins de Fer*, February, July and October, p. 44.

Kolm, Serge-Christophe (1968), "La théorie économique générale de l'encombrement", *Futuribles*, SEDEIS, p. 82.

Laffont, Jean-Jacques (1977), *Effets externes et théorie économique*, Monographies du séminaire d'économétrie, No. 13, éditions du CNRS, Paris, 200 pp.

Laffont, Jean-Jacques and Jean Tirole (1994), *A Theory of Incentives in Procurement and Regulation,* The MIT Press, Cambridge, Massachusetts, London, 705 pp.

Le Duc, M. (1995), *Services publics de réseau et Europe : Les exemples de l'Allemagne, de l'Espagne, de l'Italie et du Royaume-Uni*, Les études de La Documentation Française, Paris, 1995.

Morrison, Steven, *The Theory of Optimal Highway Pricing and Investment: Comment*, Northeastern University, Boston, Massachusetts.

Nash, C. (1993), "Developments in transport policy: rail privatisation in Britain", *Journal of Transport Economics and Policy*, Vol. XXVII, No. 3, September, pp. 317-322.

Office of the Rail Regulator, *Framework for the approval of Railtrack's track access charges for freight services*, a consultation document, October 1994, 19 pp.

Preston, J. (1996), "The economics of British Rail privatisation: an assessment", in *Transport Reviews*, Vol. 16, No. 1, pp. 1-21.

Quinet, E. (1990), *Analyse économique des transports*, PUF, Collection Économie, Paris, 302 pp.

Allais, Maurice, Del Viscovo, Duquesne de la Vinelle, Oort and Seidenfus (1965), "Options de la politique tarifaire dans les transports". Report prepared at the request of the Commission of the European Economic Community, *Studies Collection, Transport Series*, No. 1, Brussels.

Rathery, A. (1984), "De nouvelles perspectives pour la politique des infrastructures de transport dans les Communautés européennes ? Possibilités offertes par une approche pragmatique de la tarification d'usage des infrastructures", *Transports*, No. 297, October, pp. 358-365.

Rothengatter, W. (1991), "Deregulating the European Railway Industry: theoretical background and practical consequences", *Transportation Research*, Vol. 25A, No. 4, pp. 181-191.

Small, Kenneth A. (1990), "Highway Infrastructure: Crisis in Finance or Crisis in Management", *Review of the University of California Institute of Transportation Studies*, Vol. 13, No. 2, February, pp. 4-8.

Small, Kenneth A., Clifford Winston and Carol A. Evans (1989), *Road Work: A New Highway Pricing and Investment Policy*, The Brookings Institution, Washington, D.C.

The Department of Transport (1993), *Rail Freight Privatisation (the Government's proposals)*, London, 27 pp.

Vickrey, W. (1987), "Marginal- and Average-Cost Pricing", in: Eatwell *et al.* (eds.), *The New Palgrave Dictionary*, Vol. III, Macmillan, pp. 311-318.

Vickrey, W. (1994), *Public Economics* (Selected papers by William Vickrey), Edited by Richard Arnott, Kenneth Arrow, Anthony B. Atkinson and Jacques Dreze, Cambridge University Press, 558 pp.

Walters, A.A. (1987), “Congestion”, *The New Palgrave Dictionary*, pp. 570-573.

Wilkinson, A.W. (1992), “La situation politique des British Railways”, *Le Rail*, July, No. 33, p. 28-2.

FRANCE

Luc BAUMSTARK
Chargé de Mission
Commissariat Général du Plan
Paris

Alain BONNAFOUS
Professeur
Laboratoire d'Économie des Transports
Lyons
France

SUMMARY

Paris/Lyons, January 1998

1. INTRODUCTION

The rail sector very quickly came to be regarded by economists as a typical example of a “natural monopoly”. In fact, like other networks, though often even more emphatically, it displays all the characteristics which tend to compromise the theoretical pact between the mechanisms of the competitive market and the optimum allocation of resources (in the Pareto sense). On the one hand, on such a market the break-even prices lack the qualities necessary to induce the economic agents to take optimal decisions. On the other hand, it is not possible to find a satisfactory price system that enables an optimum to be decentralised on the basis of the decisions of individuals.

The sector is, in fact, characterised by increasing returns resulting from multiple indivisibilities: functional indivisibilities such as the co-ordination of activities between upstream and downstream of production, for example, between infrastructure choices and commercial policy, but also technical indivisibilities such as that determined by the continuity of the network. To this there should be added various effects such as “economies of scope” which enable the operator, by diversifying the services offered, to reduce unit costs or important network externalities, both positive and negative, or, finally, investment irreversibilities. Thus, the rail sector accumulates the conditions for the appearance of what are commonly called *market failures*. This largely explains the sector's mode of operation in the early 80s and the major role played by government throughout its history.

However, the efficiency of this monopolistic organisation and the frequent public intervention in the sector have gradually been called into question on two counts. Firstly, it has been noted that the theoretical justifications for such an organisation apply only to part of the monopoly (mainly the infrastructure). Secondly, the very real advantages afforded by such an integrated structure have been offset by its no less real disadvantages (mainly its inability to exert sufficient pressure on production costs). Thus, the policies applied in this sector have sought to eliminate these weaknesses by reintroducing, wherever possible, various kinds of competitive pressure. Infrastructure access charges have become a keystone of the reforms. This question forms the subject of the first part of our paper.

There have been many modern theoretical developments in this field and our presentation does not seek to be exhaustive in this respect. However, there are two theoretical principles that deserve closer examination and these will be the focus of our attention in the second part[1].

The first principle states that infrastructure pricing cannot be separated from investment choices. The setting of "optimal" user fees presupposes the prior definition of a certain quality of service. For example, from the quality of service required there follows an acceptable saturation level which, in its turn, determines a co-ordinated policy of investment and demand management through infrastructure user fees. Hence, the theoretical economic toll is a composite toll, comprising a cost toll relating to the costs which can be directly allocated to the users of the network and an adjustment, sometimes called a pure toll, which enables the manager to produce the necessary quality of service by adjusting, as precisely as possible, the available capacity to the expressed demand (note that this second component only concerns the part of the network with no surplus capacity).

The second principle involves relating pricing to the difficult question of covering the fixed and joint costs which within the rail system, as in many other networks, represent a large proportion of total costs. The traditional theoretical solution based on marginal cost pricing, considered optimal, was to have these costs covered by the State budget. However, the perverse effects of systematic subsidisation, nowadays decried as public failures, mean that such a solution can no longer reasonably be regarded as optimal. Nevertheless, imposing budgetary constraint on the enterprise responsible for the infrastructure is not sufficient in itself to ensure the efficiency of the system. The pricing method and the formulas for allocating the fixed costs which, by definition, cannot rationally be allocated to any particular user, have definite implications for the social surplus which the system can produce. One of the main difficulties encountered in connection with pricing is that associated with the manager's technical and economic (but also political) ability to charge differential, even discriminatory fees.

In the last part, we compare these abstract principles with actual practice.

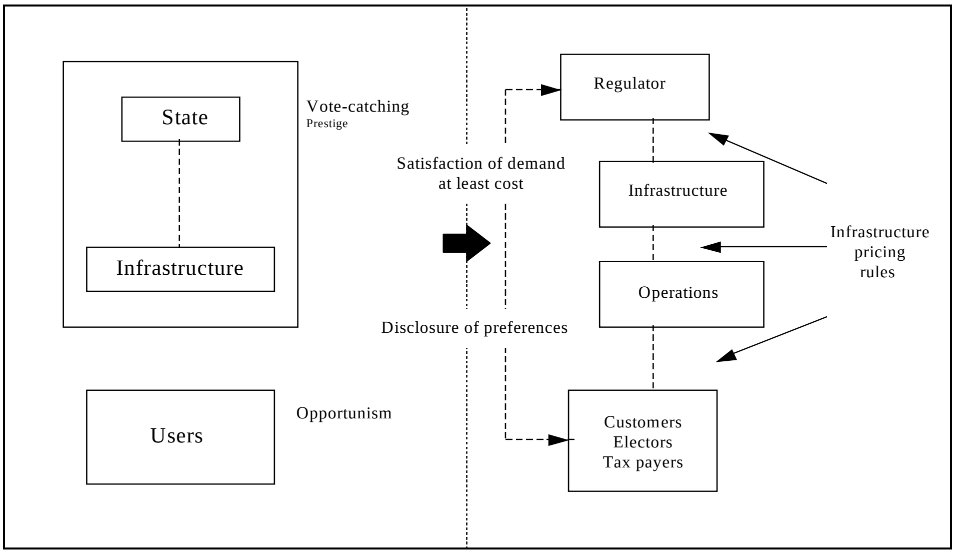
2. INFRASTRUCTURE PRICING, KEYSTONE OF INSTITUTIONAL INNOVATION

2.1. Network opening, redistribution of roles and pricing system

The idea of opening up the network is not a new one. Thus, networks have been identified in many different industries: the high-tension lines in the power distribution sector, optical fibre systems in telecommunications and delivery and collection in the postal sector, to which could be added numerous services such as the transport of electricity and information and the routing of mail. Although, as far as the network is concerned, it seems preferable to retain a monopolistic structure because of the economic characteristics of the infrastructure, this is no longer true of the services for which, on the contrary, competition seems possible.

In order to solve the problem of opening up the network, it is necessary to visualise how certain decisions which were taken internally within a single institution can be made the subject of commercial negotiations. The analysis must therefore focus on the types of relations that may exist between the decisionmakers, the private or public operators and the users. The configuration of the system and the actors is illustrated by the diagram reproduced in Figure 1.

Figure 1. **A traditional organisation chart**



It is thought that, under certain conditions, the intrusion of economic mechanisms into the integrated model will introduce transparent processes, thus obliging the various actors to disclose their preferences. This then leads to the explosion of an extreme situation in which the State produces the infrastructure regardless (or almost regardless) of the demand, of which it has only a faint idea. In this situation, the user benefits from a good without knowing its cost.

This can be improved upon by identifying intermediate levels. It is possible to distinguish between the production of the infrastructure and its management. Attention may also be focused on the behaviour of the agents and its determinants: the efficiency of the system as a whole depends on the efficiency of the means deployed to encourage and promote partnership between the economic actors and the authorities at various administrative levels. In short, it is a question of organising effective modes of interaction between the public and the commercial spheres.

On the one hand, it is necessary to establish relations between the State or, more generally, the authorities responsible for organising public transport and those directly concerned with the infrastructure so as to minimise infrastructure costs. This is part of the new theory of regulation which seeks to define more precisely the contractual framework within which these relations are to fit.

On the other hand, it is necessary to make sure that all the transport operators have access to the infrastructure on fair terms. This poses a specific problem with respect to the relations to be established between the infrastructure operator, public or private and the downstream service providers. Should or should not the monopoly be authorised to supply the market downstream? Can the sole and final responsibility for the allocation of timetable slots be entrusted to a body which itself also operates the infrastructure?

The determination of the prices at which the infrastructure manager opens up his network is one of the major considerations for the sector as a whole. These access charges are important since they determine the terms of competition between modes of transport. They must also ensure the overall efficiency of the system, that is, be sufficiently high for the upstream monopoly to be able to offer a suitably maintained infrastructure where the rail system is pertinent, but not so high as to bar the arrival of new entrants. More generally, pricing policy makes it possible, on the one hand, to encourage users to programme their services better (in the sense of making the best possible use of capacities) and, on the other, to steer investment towards ensuring that the system is developed and, in particular, the network expanded, so as to give the best possible cost-benefit ratio.

When he wishes to formalise the importance of the pricing system in the resource allocation process, the economist generally analyses the consequences of price distortion. Whatever the reasons, legitimate or not, for the distortion, the consequent loss of global surplus needs to be estimated. Within this theoretical framework, assuming an initial situation in which the pricing system is optimal, any change in prices not justified by a change in the cost of the production factors or by a sudden imbalance between supply and demand would inevitably lead to the misallocation of resources, a poor investment policy and, ultimately, less user satisfaction. In this case, the surplus available in the system would no longer be maximised.

This theoretical approach, even though it clearly relies on mechanisms less complex than they are in reality, provides valuable benchmarks for the orientation of supply and ensures cohesion between different strategies capable of serving the same purpose. The pricing system may then be understood as an information and co-ordination system, and if it offers a certain number of minimum qualities, the regulation of the system, being based to a greater extent on the rationality of the economic agents, will gain in efficiency.

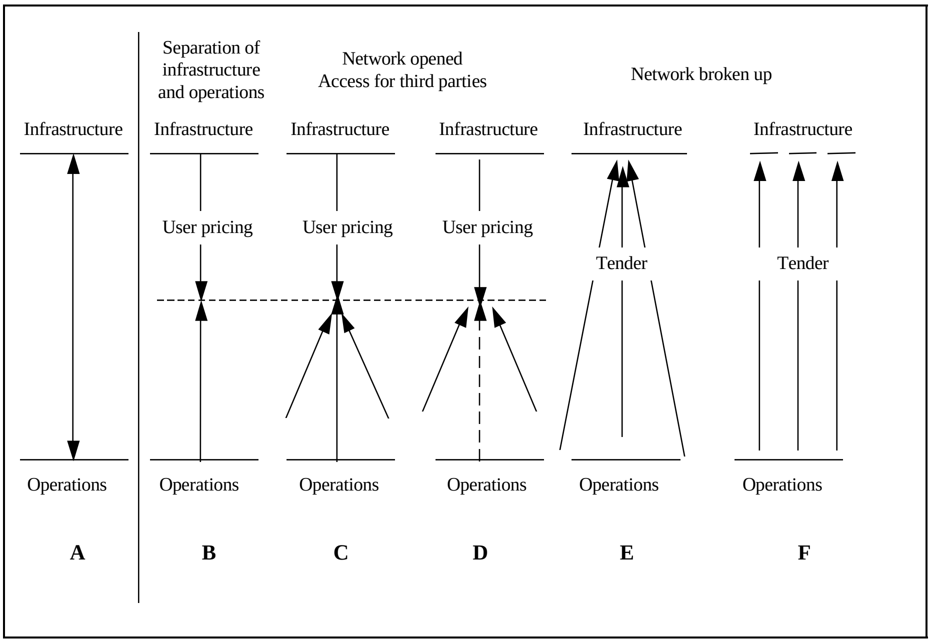
2.2. Degree of vertical separation in the rail sector and the question of access charges and user fees

These theoretical analyses underlie many of the recent trends in the European rail sector. They form the context for the adoption of European Directive 91/440, which first provided the impetus for the separation of infrastructure from operations. The Council of Ministers took a further step by issuing two new Directives 95/18 and 95/19. Before the network can be opened up, the conditions of network access must be established: licensing system, procedures for the allocation of existing capacities, coverage of infrastructure costs, network development, appeal procedures, etc. However, this Community drive is based on the reforms which many countries have already started to introduce. Although these reforms may employ different means, they apply the same basic principles, namely, the principles of decentralisation and the organisation of competitive pressure.

In practice, these widely introduced vertical separation policies are very diverse. This diversity contrasts sharply with a very homogeneous theoretical discussion of the question[2]. The main difference between all these systems appears to relate to the role of the State. Sometimes, though this is not the most frequent case, the State more or less organises the service, whereas the infrastructure is private[3]; in another configuration, the service may form the subject of a private monopoly, whereas the infrastructure remains in public hands (as in New Zealand). Under a third scenario, private companies operate a network in "partnership" with another private partner who owns the infrastructure (as in the United Kingdom).

The infrastructure user fee structure and its possible impact on the markets cannot be understood without taking into account the institutional configuration governing the management and operation of the network. A summary typology of the possible configurations is proposed in Figure 2.

Figure 2. **Typology of institutional configurations**



Situation (A) is the one that has been most frequently inherited from the past, namely, an integrated network occupying a monopoly position. As we have seen, adaptation, which may go as far as radical reform, involves creating levels of responsibility that give rise to commercial partnerships and to the introduction, with varying degrees of firmness of resolve, of the machinery of competition. Depending on the extent to which the network is broken down, several options are possible.

Assuming that the integrated monopoly is maintained, it is possible to imagine the introduction of competitive mechanisms similar to those in effect for certain urban public transport systems: invitations to tender for the management of the entire system lead to the appointment of a manager for a specified period. The integrated monopoly may be completely preserved (E). However, the network may also be broken up and tenders invited for each segment (F). The network may be divided up on a geographical or a sectoral basis, for example, with a network specialising in freight. In these various cases, the question of access charges does not really arise since the integrated monopoly is preserved: the bargaining takes place within the enterprise itself without involving any market services. The authorities may face other regulatory problems, but they will no longer have to deal directly with the question of access charges.

This does not apply to configurations (B), (C) and (D): although the integrity of the system as a whole is maintained (the infrastructure manager has authority over the entire network), the relation between the infrastructure and the operator or operators becomes distended. There are still many possibilities for reform, depending on the degree of separation between the infrastructure manager and the operator. Here the competitive pressure is exerted (with varying degrees of firmness) at the level of the infrastructure user by the (more or less) explicit means of diversification and dynamisation of the supply of services. However, in all three cases, it is a question of sending the operators a price signal calculated to influence their demand for the use of the infrastructure. The question of user fees will then be the determining factor.

Case (B) illustrates the lowest level of separation: third parties are denied access. Here, separation is intended simply as a means of clarifying the enterprise's accounts by dividing the expenditure items more clearly between those that relate to investment and maintenance costs and those that relate to operations. This model corresponds to the situation of those countries which, following the Community decision, preferred not to call into question either the integrity or the unity of their monopoly. Thus, the whole of the network continues to be run by a single operator. However, this system seems rather unstable and can be better understood as a staging post on the road to a more radical mode of organisation in which, separation having been made effective, third-party access can be encouraged. In the first case, pricing can remain a sort of accounting device by means of which the carrier is supposed to remunerate whoever undertakes to maintain and develop the infrastructure, that is to say, himself. Thus, the pricing rules are not really market mechanisms within which the parties endeavour to optimise their results. In cases (C) and (D), which are of more particular interest, the situation is quite different.

In case (C), the opening up to third parties is explicit but there remains a principal operator. Thus, the opening is still marginal and may be restricted to certain market segments, for example, the international market. This transitional situation may evolve into the much more competitive situation (D) in which, third-party access having been generalised, the historical operator is one actor among others. These two models lie at the heart of the two main scenarios for the implementation of the reforms undertaken in Europe.

We note, however, that the choice of system (C) does not protect the network concerned from sliding towards system (D). There will be de facto involvement of the international services and their demand for track will have to be reconciled with that of the national operator. Here, Community policy has clearly taken the same path as led road freight transport from progressive liberalisation of international transport to complete freedom of cabotage.

Consequently, the opening up of the network, however partial, raises the problem of how to regulate between the operator already in place and other potential operators. Even though the slot allocation rules may play a very important part, in the long run the calculation of access charges should prove decisive. Suppose, for example, that the allocation rules favoured the operator in place, say, through a practice involving “grandfather clauses” comparable to that observed in air transport, but that, at the same time, there was a heavy demand for slots, for international traffic, for example. The infrastructure monopoly could then take advantage of this excess of demand over available capacity by increasing the user price where the supply was scarce. If the use of the “grandfather clause” were too expensive for the historical operator, he could waive it and review his operating schedules in order to fall back on slots in better supply and the available capacity would then be sufficient. If it were still profitable to exercise the right despite the higher price, this would bring the infrastructure monopoly a return sufficient to finance the necessary investment in capacity.

In practice, this configuration might be relatively rare, but a similar problem might arise, including in the case of configuration (B), if the historical operator were to make, within the same space-time frame, requests for slots on behalf of different services (national and regional passenger services, freight services, etc.). If these requests were sufficiently well justified because of corresponding traffic demand, that would reduce to the expression of a need for investment in capacity on the sections in question. The historical operator might then have an interest in sending a "price signal" reflecting the shortage of available slots to each of the requesting services. It would be up to the latter either to buy because they were able to cover the cost or to transfer the demand to less busy routes or periods. This comes back to the above-mentioned choice between managing the demand and being able to finance additional capacity.

This alternative reduces to an infrastructure pricing principle which is well known but worth recalling in order properly to explore its conditions of application to rail transport.

3. INFRASTRUCTURE PRICING AND MAXIMUM EXTRACTION OF AVAILABLE SURPLUS

Infrastructure pricing may have more than one objective. The following presentation is founded on the idea that the basic principles should be established as a function of the collective efficiency of the system. Indeed, it would seem difficult to defend a pricing system which regularly deviated from that objective. Thus, this theoretical look at the problem suggests that there should be no segmenting of pricing between different, possibly contradictory objectives, namely, optimum use of the infrastructure, on the one hand, and the financing of its renewal or the pursuit of social or environmental objectives on the other. In fact, all these objectives can be reduced to the general objective of maximising the global surplus generated by the system.

On what principles should pricing be based in order to extract the maximum surplus from the system for sharing out among the various actors? This surplus can be identified only by relating it to the costs of the system and, in particular, the investment costs.

3.1. The concept of marginal cost and the optimisation of investment

Maximum efficiency obtained by marginal cost pricing in sectors with increasing returns is a decisive conclusion contributed by the theory of welfare economics. However, this theoretical result presents a number of problems. Apart from the undoubted difficulties of application, the concept is also open to theoretical objections.

In many cases, however, in discussing this issue, confusion is created insofar as the concept of marginal cost that is criticised is very often reduced to a user cost[4]. At first glance, this is hardly surprising. Understandably, since the marginal cost enters into the equation as the derivative of a total cost function, any cost factor that does not vary with production will vanish in the mathematical operation.

This classical presentation, without actually being false, can lead to inaccuracies with dangerous theoretical and practical consequences, simply because in this approach it is assumed that the investment is given and realised. In a manner of speaking, the act of investment is ignored as an optimising tool, on the same basis as pricing. However, the theoretical framework of the optimal allocation of resources cannot be satisfied with this reduction or with the solution consisting in transposing this short-term static model to the long term and treating the fixed costs as variables, thus causing them to vanish.

Indeed, taking the fixed costs into account presupposes an analytical approach quite different from that which consists in making them vanish. The fixed costs form part of the irreducible and fundamental linkage between the short and the long term. Leaving out that linkage amounts to evading the embarrassing questions raised by coverage of the costs thus incurred over several periods. From the standpoint of cost minimisation, it is well known that it is always possible to distinguish between directly avoidable costs[5], costs that can be gradually absorbed and, finally, totally unavoidable costs. This distinction then makes it necessary to bring other concepts into play.

In particular, it is necessary to introduce a terminological distinction in order to avoid confusion. The variable marginal cost is a partial cost clearly distinguishable from the marginal cost, which has a different theoretical content: the partial cost relates specifically to a short-term situation in which only the variable costs are considered, while the concept of marginal cost relates to a situation in which it is assumed that the authorities meet the demand, if need be by carrying out a project that creates a discounted surplus greater than the total of the costs necessary to produce it. Thus, as many authors have stressed[6], this concept is fundamentally linked with the investment decision.

Taking the long term into account requires the use of a programme situated upstream of the decision to invest. The fixed costs are not, strictly speaking, variabilised; however, the transport infrastructure is regarded not as a natural resource made available but as a good, to produce which the community must sacrifice resources.

Suppose that there are several technologies capable of meeting a given level of demand. Each of these can be characterised by a fixed cost, independent of the level of utilisation of the infrastructure and by a variable cost. The producer or the authorities have a certain number of possible solutions available to meet the demand. To each investment there corresponds a particular total cost function which depends on two variables, the investment I 1 and the level of production q:

$$C(I,q) = a(I) + b(I,q) \quad 2$$

Thus it is possible to distinguish between two marginal costs:

-- the short-term marginal cost when q varies for a given investment level:

$$\gamma = \frac{\partial C(I,q)}{\partial q} \quad 3 \quad ;$$

-- the long-term marginal cost when the investment varies:

$$\Gamma = \frac{\partial C(I,q)}{\partial I} .4$$

It is reasonable to assume that the producer (or the authorities) wishes to produce the good or service in question at the lowest possible price, if only to maximise the collective surplus. Thus, for each anticipated level of production, he will adopt the technology calculated to supply that good or service at the lowest price. Therefore, to each level of production there corresponds a minimum cost which itself relates to a particular total cost function belonging to the family of $C(I_i)$*5*. If it is assumed that I can vary continuously, the totality of these optimal points, which minimise the cost of production, will form a so-called long-term cost curve. We note that the curve obtained as a result of such a theoretical projection will be completely virtual. It defines a boundary of maximum efficiency of the production system at a given instant of time. Rather than an objective to be achieved, it is primarily a theoretical benchmark. It is a short-term curve envelope. This curve defines a functional relation between the long-term total cost and the quantity produced, on condition that each level of production is obtained with plant of the optimum size.

If it is assumed that the producer is able, at any time, to adapt his investment to the level of demand, there will be an infinity of short-term total cost functions. The system may be called upon to display its adaptability at any time. In this case, there is no contradiction between the short-term and long-term marginal costs. The short-term marginal costs are always defined, while the size of the plant is optimal for the production level in question.

To discard the hypothesis of permanent adaptability (divisibility) is to accept that the demand can be met by a technology or investment which is not necessarily optimal. In this situation, which is generally that of transport systems, there is almost inevitably a difference between the short-term and long-term marginal costs. There will then be several possibilities.

In the first case, that of underinvestment, the technology employed is not optimal since the short-term marginal cost is higher than the long-term marginal cost. In the second case, that of overinvestment, the short-term marginal cost is lower than the long-term marginal cost; the demand is met by an investment which costs too much for the use which is made of it. When, exceptionally, the system is optimised, the supply is precisely adapted to the demand, the two marginal costs are equal and, in this case, the marginal cost involves, over and above the short-term marginal expenditure, the optimal variations in the cost of the plant necessary to meet the demand.

> *"Under these conditions, and for practical applications of the theory of social returns, it is necessary to define the marginal cost of a specific service as the additional costs of all kinds (labour, energy, raw materials, depreciation, interest charges) resulting from the provision of an additional unit of that service when the existing fixed plant is precisely adapted to the volume of production in question[7]."*

Accordingly, the marginal cost to which pricing theory relates is defined at economic equilibrium and, consequently, with total utilisation of the fixed plant. This leads to a radical distinction between the marginal cost as described above and the partial cost. Thus, Marcel Boiteux considers it necessary, despite everything that may have been written on this subject during the last thirty years, to denounce, once again, the frequent identifying in common parlance of the marginal with the variable cost. He considers this "faulty" identification to be still "a major source of misunderstanding"[8].

This also means that the determination of the marginal costs has real economic significance only on the assumption that the infrastructure is optimally managed. For the concept to be pertinent, the maintenance operations and the investment renewal operations must be carried out at the optimal time and their cost must be minimised.

If the marginal costs thus defined are used for pricing purposes, the system will be coherent and optimal. This means that the consumers' choices will be optimally oriented since they will be encouraged to choose, among several ways of satisfying their needs within a given framework of constraints, that which is least expensive for the community. However, at the same time, the community must ensure that the demand thus expressed is always satisfied at the least possible cost.

This clarification is a necessary preliminary to the application of practical pricing tools. Within this context it is now clear that marginal cost pricing is future-oriented pricing. Accordingly, it should concern itself not with the previous cost of the infrastructure but with the use that the users will make of it. A knowledge of the users' preferences then becomes one of the keys to the global optimisation of the system from the standpoint of both economic calculation and pricing.

However, the balance between supply and demand thus achieved may give rise to very different situations on the rail system depending on the level at which the service quality is to be pitched. We have previously used expressions such as "meeting the demand" or "plant of optimum size". For example, should there be perfectly regular timetables with zero delays or would a reasonable amount of delay be acceptable if it permitted the more intensive use of capacity? Clearly, these questions prompt us to reflect on the quality of service to be provided which, in its turn, is also subject to a trade-off between costs and benefits.

3.2. The supply-demand equation or how to define an optimal quality of service

To deal with this problem, public economics traditionally relies on the notion of a mixed collective good, i.e. a good such that the quantity consumed can be distributed among the individual consumers (the good is therefore divisible): the consumption of a good by a user cannot be consumed by another, whereas certain other so-called quality characteristics remain indivisible because they concern all users. These mixed collective goods are subject to so-called congestion effects[9]. In the case of rail infrastructure, the indivisibility of its use can never be ensured inasmuch as two trains cannot share the same "slot" at the same time. The slots are shared between the different users. The use of infrastructure, from this point of view, is divisible. On the other hand, the quality of service, which depends on the reliability of the train timetables on a congested part of the network, is the same for all the users at a given instant[10]. This collective good is called mixed because it has two fundamentally different characteristics, the first -- access to infrastructure -- being divisible, the other not. Congestion is a particular case of an external effect *where the reasons why people cause and suffer it are linked with the consumption of the same service.*

In order to optimise the system and achieve the maximum available surplus on the infrastructure, the authorities must both determine the optimal level of investment and, through good pricing, manage the level of demand. Investment and level of demand are the two factors of the quality of service offered which different users value in varying degrees.

In order to isolate the problem posed by this linkage, let us consider an infrastructure, the cost of using which is very low and may therefore be deemed negligible[11].

Let us suppose that all the users *i* liable to use the infrastructure place different values on the use of this good, for a given quality of service. We then obtain an inverse demand curve which we will denote by $P_i(Q_i, g)$, where Q_i represents the utilisation of the infrastructure by the individual *i,* and *g* an indicator of service quality.

The quality of service offered on this infrastructure depends on two parameters: the number of users on the infrastructure and its physical characteristics, which depend on the level of investment *k*. This level of investment varies with time as a function of the outlay made by the operator, $\psi(k)$ 6.

Taking into consideration several time periods t, we write $P_{it}(Q_{it},g_t)$, with $g_t = (Q_{it},...,Q_{nt},k)$ the quality of service at period t. We then suppose that each user integrates this quality of service, $P_{it}(Q_{it},g_t)$ being the value that the user i accords to the use of the infrastructure at a given level of quality g. This value may evolve over time.

The individual surplus is given by the difference that exists between the value that the individual attributes to the use of the infrastructure and the cost which that use represents for him, remembering that in this case the user cost, in its strictest sense, is seen as negligible. This cost is therefore limited to the deterioration in quality of service suffered by the user, due to the utilisation of the infrastructure by others. This cost is all the more important the greater the number of users and the weaker the investment. If we take D_t as the deterioration of service quality at time t and V_{it} the value that the user i gives to this deterioration, the surplus of user i for each period t is expressed by:

$$P_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right)$$

Now let us assume that the operator is seeking to maximise the social surplus which this infrastructure is capable of generating. For each individual i and for the entire length of time in question, we have:

$$\sum_{t=1}^{T} \left(\int_0^{Q_{it}} P_{it} . dQ_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) \right)$$

The global surplus is obtained by summing over all the users and subtracting the investment cost. This calculation leads to the following expression:

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \left(\int_0^{Q_i} P_{it} . dQ_{it} - Q_{it} V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k \right) \right) - \psi(k)$$

The maximisation of the surplus depends on two variables, the utilisation of the infrastructure and the level of investment.

The first n partial derivatives have the form:

$$\frac{\bar{o}S}{\partial Q_{it}} = P_{it} - V_{it} D_t \left(Q_{1t} \ldots Q_{it} \ldots Q_{nt}, k\right) - \sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\partial D_t}{\partial Q_{it}}$$

The relation for the investment is written as follows:

$$\frac{\partial S}{\partial k} = -\sum_{t=1}^{T} \sum_{i=1}^{n} V_{it} Q_{it} \frac{\partial D_t}{\partial k} - \Psi'(k)$$

7

Thus, the first-order conditions lead to the following two relations:

$$P_{it} = V_{it} D_t \left(Q_1 \ldots Q_i \ldots Q_n, k\right) + \sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\bar{o}D_t}{\partial Q_{ti}} \quad \textbf{(1)} \text{ and}$$

$$\Psi'(k) = -\sum_{t=1}^{T} \sum_{i=1}^{n} V_{it} Q_{it} \frac{\partial D_t}{\partial k} \text{ 8} \qquad \textbf{(2)}$$

Equation (1) establishes the principle of the pricing rule. The first term of this expression represents the "cost" accepted by the user *i* resulting from the loss of quality of service associated with congestion. The second term represents the value of the quality of service lost by all the users because of the last utilisation by the user *i*: this last term represents the external cost of congestion for which the latter is responsible.

Using classical pricing terminology, the optimal pricing is such that the price corresponds to the sum of the private marginal cost and the social marginal cost. The first term is already borne by the user. The toll to be applied should therefore correspond to:

$$\sum_{j=1 et j \neq i}^{n} Q_{jt} V_{jt} \frac{\bar{o} D_t}{\partial Q_{ti}} \text{ 9}$$

If we now consider the second equation (2), which incorporates the investment dimension, it indicates that the capacity should be developed until the marginal investment is equivalent to all the congestion costs avoided.

Finally, and this is crucial, the optimum in terms of social efficiency will only be ensured if these two conditions are satisfied. The golden rule of marginal cost pricing consists of these two conditions, which theory suggests should be kept together. In common parlance, it could be said that this dual logic consists in establishing a price which, wherever there is saturation of the network, ensures either that demand is sufficiently well managed for the available capacity to be still sufficient or that the cost of the necessary investment in capacity can be covered. However, the quality of service production function may have different characteristics depending on the importance accorded to the intensity of the demand. The optimisation procedures will then be more or less determined by scarcity phenomena.

The congestion toll described above may take several forms, depending on the degree of divisibility of infrastructure use. The less the divisibility, as in the case of road transport, the more the intensity of use will affect the quality of service. But the greater the divisibility the less perceptible this effect and the more pricing should be oriented towards scarcity management systems, as in the case of car parks. The analysis should therefore be focused on the technical and economic relationship between the infrastructure and the uses of that infrastructure.

The rail system does not readily lend itself to this type of analysis. The existence of timetable slots seems to give credence to the idea that the use of the infrastructure is totally divisible and that there is therefore a strong user rivalry among the various consumers. Thus, it is theoretically possible to establish a market in user rights, for example, by auctioning slots, and thus reconcile supply and demand. It seems that this might be practicable on railway lines dedicated to similar kinds of traffic since, in this case, the slots auctioned would tend to be homogeneous. However, the only configuration that would seem to lend itself to this exercise is that of lines specialising in freight and open to several carriers, which could only apply to the lines, still to be organised, of a trans-European rail freight network. The first “freight corridors” to be established are not organised along these competitive lines.

In most cases, competition remains latent and overshadowed by slot allocation rules which predate the reforms. However, it is very real wherever network saturation is a problem, which brings us to the difficult question of capacity.

Rail capacity on part of the network can be very roughly and provisionally defined as the maximum possible number of movements that can be handled per hour. The reality, however, is much more complex.

On the one hand, capacity is determined by the characteristics of the infrastructure itself: the configuration of the lines and sets of tracks, the existence of community lines, switches, track intersections, the speed permitted by the design of the track, gradients, etc. Moreover, the capacity also depends on the utilisation of the infrastructure: type of trains (length, axle load, etc.), speed, number of stops, etc. The train schedule is then decisive and its organisation will have an impact on the network's effective capacity, which will be all the more considerable the more heterogeneous the slots concerned. The capacity will then depend on the train schedule as it has been drawn up. The scheduler and the rules of arbitration on which he relies will thus play an important role, a role which, in railway tradition, has not been much influenced by the considerations of the economist who, for his part, is keen to maximise the surplus that can be extracted from the infrastructure. To that end, he analyses the economic advantages which the consumers derive from using the infrastructure. For example, a decrease in the time taken by a train to travel between two stations may lead to a reduction in the technical capacity of the infrastructure by eliminating slots, but that reduction in capacity may be justified because it results in a net gain in surplus production. Conversely, the economist lacks mastery of the complex relations between the characteristics of the infrastructure, the possible operating programmes and the response of the final demand.

This difficulty involves both infrastructure pricing and project evaluation. Whether it is a question of calculating a developing marginal cost or the return on an increase in capacity, it should be kept clearly in mind that such an increase can very often be achieved in different ways: an additional junction, flyovers for avoiding track intersections, higher speed limits, passing track en route or in the station, improved block arrangement, better power supply, safety and traffic control installations, restoration of alternative routes, etc. As with the roads, the optimum infrastructure capacity is not a technical but an economic factor which, however, for rail forms part of a complex universe of technical solutions.

Here, we have one of the explanations for that common tendency of railway reforms which “separate” infrastructure management from transport production. The unified monopoly operator or separate operators not being omniscient, whatever the theoretical optimisation models may suppose, have not efforts been made to improve the efficiency of the system by giving the carrier the opportunity to make the best possible use of the signals sent him by the infrastructure manager and vice versa? Thus, it is for the carrier to interpret the infrastructure user price signal and for the infrastructure manager to interpret the demand signals he receives from the carrier. These “interpretations” could be based on the teachings of public economics.

However, to the above-mentioned difficulties we should add another associated with a familiar feature characteristic of most rail systems, namely, the presence of very considerable fixed costs and modest marginal costs over most of the network, i.e. wherever there are no saturation effects. The principles we have noted then lead to pricing which can result in a considerable deficit which it is the community's responsibility to meet. Whence the now acknowledged need to take this “budgetary constraint” into account.

3.3. Allowing for the budgetary constraint

The doctrine according to which the producer price of industries with diminishing costs should depend only on the marginal operating costs, and the rule according to which the authorities should cover all the fixed costs from taxes, first made an explicit appearance in the railway literature of the late nineteenth century. The French tradition of the economist engineers of the École des Ponts et Chaussées made a big contribution to these developments and their application. The debate became more animated following the appearance of Hotelling's article[12] in 1938. He concluded that the deficit resulting from the application of this global pricing principle should be financed by global taxes which, like the taxes on rent charges or inherited income, are supposed, in theory, not to affect the marginal behaviour of the economic agents.

Nevertheless, there may be a conflict between this theoretical viewpoint and another scarcity phenomenon, the public finances. Moreover, it assumes that resources are optimally allocated by the public operators, which is not necessarily the case when they believe themselves to be free of all financial constraints. It is to correct these dysfunctions that an entire theoretical school has devoted itself to justifying the addition of a budgetary constraint to marginal cost pricing. There are several ways of understanding this approach.

For some, this concern for a balanced budget is, in the words of Serge-Christophe Kolm, no more than the obsession of *"a short-sighted and narrow-minded accountant, who knows nothing of economics, trying mistakenly to transpose the criteria of the private to the public sector*[13]*."* By refusing to consider the main analytical conclusions of welfare economics, the supporters of the balanced budget deny themselves the possibility of satisfying, with the tools of economics, the demands of the public interest. Without rebutting this position directly, Mark Blaug has made it the target of some equally critical remarks which reflect the difficulties created by this pricing principle. Thus, according to Blaug, what the Anglo-Saxon authors call the French school[14] has trouble in taking the existence of deficits into account: *"The characteristic feature of the French contributions to the literature (on marginal cost pricing) is a total inability to take into account the problem of deficits in the diminishing cost industries which, indeed, hardly receives a mention*[15]*."*

For others, there are fewer disadvantages (i.e. less loss of global surplus) in distorting optimal pricing to control the deficit than in leaving the latter to drift while seeking to adhere strictly to the optimal pricing principle. Thus, in France and Europe, there has been a slow swing of the pendulum: the arguments in favour of marginal cost pricing no longer convince the authorities, who are more concerned about the financial situation of the public corporations, especially that of the railway companies which are steadily losing modal share, as well as about the unfortunate effects which the systematic covering of the deficits is having on the management of those enterprises.

Driven by the structural difficulties of the public finances, this trend is also based on the failure of a pricing system which requires transfers between taxpayers and users to ensure that the cost of the service to the community is explicitly weighed against the interests of those who use it. The gap between those who benefit from the system and those who finance it leaves a space within which the economic agents can conceal their preferences. The deficit subsidies may also hide and hence permit inefficient operation. Now, if there were no strong incentive to seek the minimum average cost, the willingness of the State to close any gap between the marginal cost and the average cost would result in enormous waste, which would be all the more enormous as it would probably be invisible and almost undetectable[16]. Thus the balanced budget constraint is aimed at an efficiency deficit which goes far beyond the traditional criticisms levelled against the champions of marginal cost pricing concerning the difficulties of evaluating the marginal cost and the technical, political and institutional barriers to implementation.

The following list of criticisms, without being exhaustive, will serve as an illustration: the pronounced indivisibilities of the infrastructure would lead to "sawtooth" pricing incompatible with long-term decisionmaking by the economic agents; the lack of rules for allocating certain cost elements would make competition impossible because of the existence of cross-subsidies; differential pricing on the network would bring into question an entire spatial standardization system; the practice would make it necessary for both users and authorities to gather, at great expense, information on the competitive structure of the market, on externalities, on the elasticity of the demand, etc. and, in short, the additional costs which the public would have to bear to implement these pricing systems would be out of all proportion to the advantages which they are supposed to bring.

As Vickrey points out, we are constantly on the horns of a dilemma from which it is difficult to escape completely. On the one hand, theoretically, the application of marginal cost pricing ensures that infrastructure utilisation is optimised but, considering the financial scarcity constraint, we then deprive ourselves of information on the real value of a new project or, if the project has already been carried out, about whether it is still worth operating. On the other hand, with the balanced budget constraint, we can be sure that the project or its operation are worthwhile, but we do not know whether the infrastructure is being utilised to best advantage[17]. More generally, by employing crude regulatory mechanisms, by excluding a number of users and by eliminating certain activities, the application of such a rule might lead to a serious loss of social efficiency[18.]

All the solutions proposed for introducing a budgetary constraint into the pricing system must face this radical criticism: the allocation of the fixed costs to a particular user or a particular use remains largely arbitrary. Thus, the calculations form the subject of endless discussions about the relevance of a particular distribution scheme. It might seem that subtle cost accounting could reduce the proportion of non-apportionable costs and thus eliminate the problem. However, these methods can always be debated and offer no solution for a usually still significant residuum.

Now, from the standpoint of the optimal allocation of resources, the pricing system should not be mainly concerned with distributing the costs but, more fundamentally, should favour the achievement of a surplus. The objective is much more ambitious. Leaving aside the utilisation costs, for which the allocation procedures do not, in theory, pose any particular problems, it is clearly the availability of the infrastructure that must be reflected in the scales and not its effective consumption. Then, the efficient mode of contribution is not to seek to allocate costs but to find a means of extracting the surplus needed to finance the infrastructure while ensuring that the surplus is achieved. From the moment that this surplus exists (it is the role of public economic analysis to locate it), there exists a pricing system to bring it to light.

It is the segmentation of the demand which, in this respect, makes a vital contribution to economically efficient pricing. The theory of surpluses leads to precedence being given to pricing systems based on this latter principle. In practice, it may lead to the adoption of the monopolists' rule to the effect that the best pricing consists in imposing the charges that the traffic will bear.

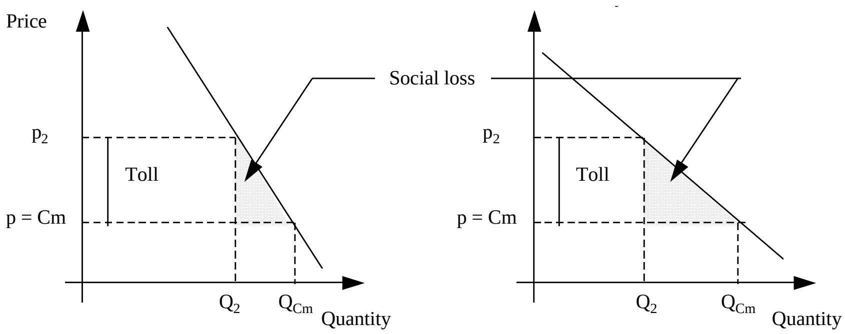
In fact, the question of the social loss involved in pricing that deviates from the marginal cost, like the question of the deficit and the limitation of its perverse effects, makes sense only if price unity is assumed. As soon as a discriminatory approach is adopted, the difficulty may disappear. Price discrimination, which takes into account the response to prices of each segment of demand, then makes it possible to increase the global surplus since the number of users will increase, while ensuring better coverage of the costs for the producer. The introduction of socially efficient price discrimination turns pricing completely upside down. It should be distinguished from the pricing principle, often encountered in the literature, that everyone pays his share, which makes even less sense inasmuch as the allocation of some costs is arbitrary. It forms part of another approach which authorises any use of a good calculated to generate a positive net surplus. Under this condition, not only will any discriminatory pricing system be neutral from the standpoint of the optimal allocation of resources, but it will be totally justified from the standpoint of the community since it will enable a surplus to be generated. Thus, the first principle considered to characterise optimal pricing gives way to another principle.

A pricing system is deemed to be economically satisfactory if the operator procures for each user a share in the absolute utility of the service sufficient to constitute an effective incentive for him to use the infrastructure. The contribution to this service is then said to be fair as long as, for each consumer and for each use, it does not exceed the net value of the utility he derives from it which, it should be recalled, must be positive. Although often disputed[19], this approach, based on tapping into a surplus, has a definite advantage in relation to the problem of financing which, as Jules Dupuit suggested, consists in "demanding as the price of the service provided not what it costs the provider but an amount commensurable with the importance attached to it by the one for whom it is provided[20].

This basic principle remains very theoretical and there are major difficulties to be overcome before it can be applied. However, the tools do exist. In 1956, Marcel Boiteux[21] proposed a solution which marked an epoch in the history of economic thought. The literature has even associated the name of the author with that solution so that it is customary to speak of Ramsey[22]-Boiteux pricing. Anglo-Saxon authors also refer to this principle as the "Inverse Elasticity Rule". Tradition has it that this seminal article offers a general solution of the problem of the production and the Pareto-optimal pricing of a public monopoly obliged to balance its budget within the context of a competitive economy.

Here, then, the decisive concept of elasticity of demand makes its appearance. Thus, if the enterprise is considered to have several markets at its disposal, observance of the budgetary rule which requires deviation from the marginal cost will lead to the prices for each market being so determined as to make the mark-up between the consumer price and the marginal cost proportional to the inverse of the price elasticity. In practice, this comes down to saddling the goods or services for which the demand is relatively inelastic with a larger share of the deficit. This pricing practice harks back intuitively to a mechanism well known in the field of taxation: the loss for the community resulting from the imposition of a tax (that is, in this case, the reduced consumption of a good consequent upon an increase in price) will be the greater the more elastic the demand.

Figure 3. **Social loss and elasticity of demand**



Assuming a monopolistic market on which the demand is very elastic, the slightest variation in price on this market will be reflected in substantial changes in the pattern of consumption. Thus, the so-called Boiteux rule consists in taking advantage of different relative market situations. The steeper the slope of the demand curve on a market, the lower the elasticity and the more limited the social loss resulting from a deviation from the marginal cost. Conversely, on a highly elastic market, the slope of the curve will be flatter. In this case, a deviation of the price from the marginal cost will be reflected in a heavy social loss.

Thus, when by necessity an enterprise has to cover the whole of its costs and hence, in the case in question, when it is obliged to deviate from the marginal cost, this theoretical demonstration makes it possible to justify placing the strain on the consumption which is the most inelastic. When the demand is relatively inelastic, the deviation of the Ramsey-Boiteux price from the marginal cost will be small and hence the deviation at the optimum will be minimised. The mark-ups will be greatest on the least sensitive demand. Thus, this method of pricing seeks not to distort the price signal sent to the most sensitive users in order that they may not significantly modify their pattern of consumption and to levy the charges on the less sensitive users who will not reduce their consumption more than slightly relative to the social optimum, even if the prices are raised.

Thus, returning to the allocation of non-apportionable fixed costs, where a monopoly can rely on several products it should parcel out its fixed costs according to the sensitivity of the demand. Prices are raised sharply where the demand is not sensitive and reduced where it is. The quantity consumed on each market remains as close as possible to the consumption which would have been observed in the first-best case. The optimum obtained maximises the social surplus subject to the constraint of a balanced company budget or, if this constraint seems too harsh or inaccessible, by assigning in the Ramsey-Boiteux optimisation programme a scarcity coefficient which overestimates the collective utility of the public contribution.

As Boiteux has himself been pointing out since 1956[23], to the difficulties of application of this pricing method there must be added the unrealistic assumptions of the model. At the same time, the author questions whether the practical application of the results obtained is of any real interest. In fact, the application of this rule poses a number of problems, in particular, by requiring a knowledge of demand elasticities. More generally, the necessary hypothesis of an omniscient economic supremo is obviously unrealistic. The fact that it is a question of pricing the use of the rail infrastructure and not the transport service itself is a further complication. The carrier is then the source of the demand, in the sense that he needs slots and he himself responds to a combined demand: that of the users or shippers and that of the transport organising authorities who "purchase" qualities of service from him. Thus, the effects of infrastructure pricing on the final demand are linked with the pricing principles which the carrier himself applies. For the system to be coherent and economically efficient, it is very likely that the two operators will have to price in accordance with similar principles, but this linkage has still to be subjected to a complete analysis.

This brings us to an essential conclusion concerning the pricing system which Maurice Allais considers to be one of the key elements distinguishing his theory of the markets economy from the standard model[24]. Within this more general theoretical framework, which seeks to maximise the collective surplus, the deficit constraint can be removed provided there is no objection to questioning price unity. That is an intuition already clearly expressed by Jules Dupuit[25] in his time. If the surplus generated by an infrastructure is greater than the cost of putting it in place, then there necessarily exists a pricing system capable of tapping into this surplus to obtain the sums needed to finance the project while maintaining maximum social utility. The achievement of the surplus becomes the challenge of the pricing procedures. It can justify the transition from a logic of differentiation based on costs alone to a logic of discrimination based on segmentation of the demand. It is not a question of differentiating prices solely on a user pays basis, which makes little sense since, once again, a portion of the costs cannot be distributed in accordance with this principle. The discrimination of the demand should be based on the principle of he pays who can, especially when that is the only way of obtaining a return on projects with a positive global net utility. Otherwise, if the project is not financed out of the general budget, a type of solution now considered best avoided, it will never materialise.

We then enter a pricing universe that is more complex but still regulated by marginal cost in which it is less a matter of minimising the effect of a deficit linked to optimal pricing as of seeking to maximise the surplus by differentiating the pricing, segmenting the customer base and, finally, obliging all the economic actors to disclose their preferences. These principles now need to be compared with the actual results of implementing the most significant recent reforms.

4. APPLICATION OF PRINCIPLES AND NETWORK CONSTRAINTS

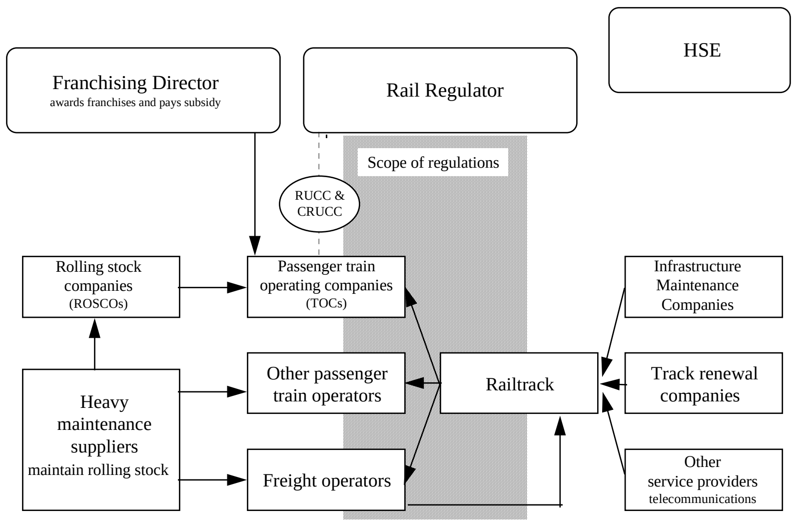
In this final part of our paper, we shall compare the few theoretical principles that can be derived from economic analysis with the practices of networks which, as a result of having been reformed, should have solved the infrastructure pricing problem. Accordingly, we shall examine, in turn, the British, German and French cases.

4.1. The British experience: making a surplus[26]

The restructuring of British Railways began in the early 80s. This long process, intended to improve the efficiency of the sector, led the authorities gradually to modify the enterprise's internal organisation. On 1 April 1994, the reforms took a further step forward with the entry into force of the Railways Act which prepared the ground for a phased privatisation of the sector by programming, in particular, for the separation of operations from infrastructure. The enterprise's transport services have thus been separated into smaller entities which, from the outset, the authorities have made clear they intend to privatise. As for the infrastructure, it has been assigned to a new, private-law company held by the State, Railtrack. This company, still a monopoly, retains operational control over the traffic, allocates capacity and, above all, is responsible for pricing use and determining the fees to be paid by the various operators to ensure that the costs are covered[27]. The company, which initially was left in public hands and even benefited from investment subsidies, was privatised in 1996. BR's passenger transport business has been split up into 25 separate entities which have been placed under private sector control by introducing a franchising system. The freight business has all been sold off to the private sector and opened up to competition. Thus, Railtrack buys services from and sells them to a range of operators.

At the same time as splitting up BR, the Railways Act established a powerful regulatory system based on three bodies. The first, the Office of the Rail Regulator (ORR), is mainly responsible for supervising infrastructure access and pricing. It establishes the rules of competition and oversees their application, especially in the interests of the customer. The task of the Office of Passenger Rail Franchising (OPRAF) is to grant franchises and supervise the correct application of the terms and conditions by each franchise holder, in particular as regards the consistency of the services actually provided. Finally, the Health and Safety Executive (HSE) ensures that the safety regulations are observed. It issues rules governing the design, construction and operation of rolling stock, infrastructure and equipment.

Figure 4. **The British reforms**



Source: ORR Report, 1996-97.

Each operator signs a track access agreement with Railtrack. Two sectors should be distinguished. Initially (7-15 years), the franchised passenger line operators have been granted the access rights necessary to provide the services stipulated in their specifications. On certain routes they enjoy protection which will eventually be withdrawn. The existing freight carriers have been granted initial rights to enable them to satisfy their present customers. Apart from the time slots allocated in connection with these rights, there are others which are open to competition.

The splitting up of the network immediately posed the problem of the allocation of costs among the various activities. It was decided that the costs should be allocated to the various sectors in such a way that each sector bore the costs of the fixed assets and personnel of which he was deemed to be the principal user. The basis for this pricing is that it must be sufficient to ensure Railtrack a certain return on its assets.

The rail operators pay Railtrack infrastructure user fees intended to cover the network utilisation and signalling costs and the cost of supplying power, where appropriate. Thus, the overall aim is to pursue a balanced-budget pricing policy which also takes into account the ability to pay of the applicants for time slots. Accordingly, it is not possible to speak of Ramsey-Boiteux pricing because of the special terms granted to the franchise operators.

In fact, the pricing system differs depending on whether one considers the operators who provide services under franchise (subsidised services) or the operators who purchase time slots on the network. The companies which operate passenger lines under licence are in a special situation since in this case the structure and the level of the access charges applicable are directly controlled by the regulatory body. The access charges applicable to the passenger lines operated under licence[28] are based on a cost allocation study that uses the concepts of avoidable[29] and additional[30] costs.

In the event of a time slot purchase, Railtrack is free to negotiate its prices, although the contract must be approved by the Regulator. The fees are negotiated but subject to approval based on the principles established by the Office of the Regulator.

The general principle requires that the fee structure should not deviate too far from the value of network access for the users and that it should enable Railtrack to recover all the costs actually incurred in connection with the transport of goods, to which there should be added a possible contribution to cover the joint costs shared with the passenger services. Thus, the pricing rule should be such that the minimum price is not less than the avoidable costs occasioned by the service concerned. The price should be less than or equal to the costs which Railtrack would incur if the operator were alone on that portion of the network and then had to assume all the costs. The price should not, when the various cost factors are taken into account, differ appreciably from one user to another.

In addition to the transparency of the charges being more difficult to achieve for freight than for passenger transport and it being difficult to determine precisely the share of the cost directly apportionable to a particular service, the allocation of the joint costs is a real headache.

The reforms have been the target of two main criticisms. Some consider that Railtrack, by setting low prices for freight transport, for which it is competing, may seek to shift the fixed costs onto the passengers and thus improperly obtain public funding to its advantage. Others, on the other hand, consider that the draconian safety regulations, though reasonable in the case of passenger transport, are less justified in the case of freight transport and, accordingly, that there has been a transfer from the freight to the passenger sector.

However, in 1995, before Railtrack was privatised, after examining the access charges applied to the franchised passenger services, the ORR concluded that they were too high and more than the operator really needed to fulfil his infrastructure renewal obligations. The Regulator imposed corresponding price reductions, thus transferring the productivity gains to the licensed operators. The charges are to be reviewed in the year 2000. Another object of criticism is the charge structure, the fixed portion being considered too high (91 per cent). This pricing is not conducive to the rational management of resources. In fact, this approach precludes the introduction into the access charge calculations of differentials in terms of rush hours or the economic value of the slots. The costs considered here do not incorporate such externalities as noise or air pollution.

When the pricing rules were being drawn up, there was a keen debate between the advocates of a commercial strategy and those who favoured a more managed approach. The arguments are important inasmuch as the same debate is now being conducted at the European level. After holding numerous consultations, the ORR concluded that it was better to place the method of calculating the infrastructure user fees on a commercial negotiation basis so as to give the operator a chance to attract all the economic agents capable of paying at least the directly apportionable costs. Thus, the main aim is to give the infrastructure manager the means to induce the maximum possible number of operators to use the network. Clearly, then, the objective is to achieve and engineer a surplus. This approach has been much criticised. The owner of the infrastructure can engage in cross-subsidisation and favour one party or another without necessarily ensuring the complete opening up of the network

4.2. The German experience: covering the costs

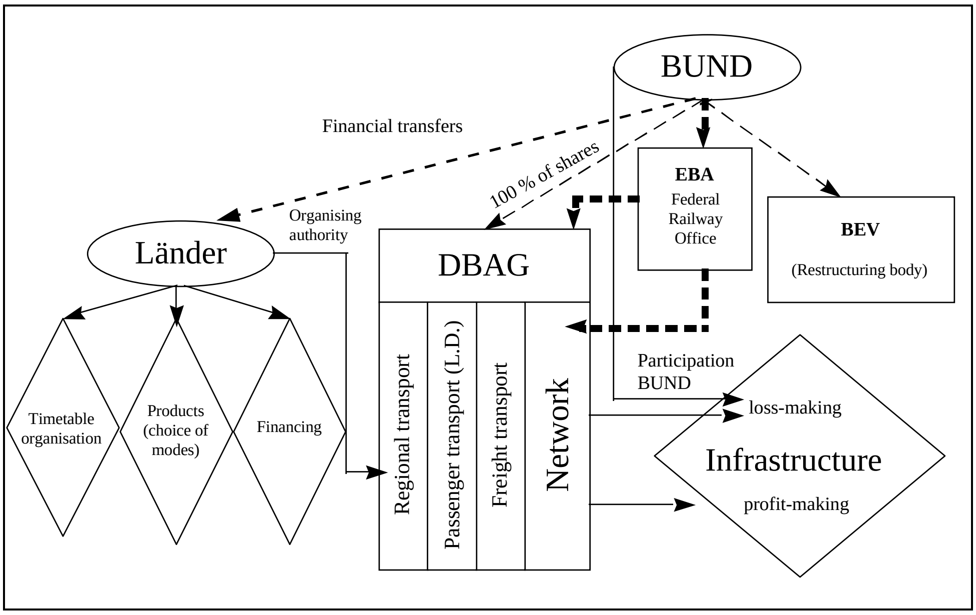
As in other countries, the amount of federal subsidies granted to the rail sector in order to balance its budget and guarantee its borrowings had become considerable while, at the same time, rail's share of total traffic was being continually eroded. Reunification only worsened the crisis and, in 1993, the enterprise (made up of the DB and the RB) had debts of more than DM 67 billion while the Government was anxious to support the development of the sector. The reform of German railways was speeded up, profoundly transforming the rail transport situation across the Rhine. It opened the way for ever keener competition on the railways, on the one hand in the short term, by placing the historical rail operator in the position of a service provider in its negotiations with each Land and, on the other, in the long term, by opening up the network to third parties. The reforms entered into effect, following an amendment to the Constitution, on 1 January 1994.

Thus, the authorities gradually moved towards a vast controlled and concerted structural reform of the railway company. The Central Government *(Bund)* set up a private joint-stock company, the DB AG, in which at present it holds all the shares but which is supposed to disappear in 1999. The rail system is organised in four independent sectors of activity: regional passenger transport, long-distance passenger transport, freight transport and infrastructure management. These sectors are eventually to be privatised. However, the Federal Government will remain the principal shareholder in the *Fahrweg* (infrastructure manager) in order to retain control over investment policy.

As in the United Kingdom, the reforms are based on new institutions. On the one hand, the Federal Office of Railways -- the *Eisenbahnbundesamt* (EBA) -- has been set up to ensure the necessary co-ordination and take care of the general missions of the railways. It authorises operations, certifies equipment and organises work on the infrastructure. It plans the work on the federal rail network, ensures that tendering procedures comply with the law, grants licences, applies investment financing agreements, commences prosecutions and settles disputes. On the other hand, the *Bundeseisenbahnvermögen* (BEV), another federal body, is responsible for clearing the debts of the former company and for administering staff and pension costs as well as financial and property charges. Central government has taken over the costs of the staff of the former DB with civil servant status, who can now be made available to the DB AG for employment under ordinary market conditions.

At the same time, to meet the increasing costs of regional rail transport, the central authorities have transferred to the *Länder*, with their agreement, the responsibility for organising and financing regional transport operations. Thus, the latter will henceforth find themselves in the position of organising authority. They will receive financial transfers from the Bund to enable them to perform this task[31].

Figure 5. **The German reforms**



Infrastructure pricing is an important component of these reforms. The Act authorises access for third parties and thus transposes into domestic law the principles of Directive 91/440. The operating subsidiaries of the DB AG pay the infrastructure manager for the use of the infrastructure. These charges are published in a catalogue which tells the operators exactly what price they will have to pay for the whole of a journey, depending on their requirements. In July 1994, DB Infrastructure published an initial price schedule applicable to all users of the network. The system was chopped up into 4 000 sections with well defined characteristics on the following principles.

The prices are identical as between the DB AG and third parties and must be identical for requests with similar requirements. The differentiation to be found in the catalogues is based on objective criteria. The prices depend on the category of line, the damage potential of the equipment used, which largely depends on the type of use, the required regularity, the volume of purchases and the length of time for which the slot is used. The basic tables are compiled on the basis of four criteria, namely, the quality of the track (essentially the permissible speed), the traffic potential (principal characteristics of the rail links) according to the type of service requested, the wear and tear (based on a variable cost analysis) and the planning quality. This last item relates to the quality of service demanded, the reliability (punctuality) indicator determining the room for manoeuvre left to the infrastructure manager. This indicator is expressed in the form of a percentage representing the margin which the DB is allowed relative to the theoretical journey time. Using these criteria, it is possible to construct the reference table shown in Table 1.

The notion of quality introduces an element of economic demand management. Here, the aim of differentiation is to define a scale of operator's requirements. In choosing a quality of service, the operators disclose their preferences. However, there are strict limits to this mechanism since the percentage is fixed for each category. It would be different if the operator could choose a level of reliability in each category.

Many weightings and modes of payment can be introduced on the basis of this table. If the stated maximum load is exceeded, the basic slot price is increased by 1 per cent for every additional 100 tons.

For trains running empty, the basic slot price for price classes P1 to P3 is reduced by 10 per cent and that for price classes P4 to P7 by 5 per cent. For engines running light, the basic slot price is reduced by 20 per cent.

Every regular slot ordered must be paid for irrespective of its utilisation. It is possible to reserve optional slots. A reservation fee of 20 per cent of the slot price is then collected. This fee is non-returnable.

Customers who order many slots are granted a price reduction which depends on the annual total of train-kilometres.

Table 1. **Quality of service reference table**

Price class	P1 High-speed traffic	P2 Express passenger traffic (main lines)	P3 Regional express passenger traffic	P4 Average-speed main- line passenger traffic	P5 Short-haul regional passenger traffic	P6 Local passenger traffic	P7 S-Bahn (urban rail transport)
Max. load (t)	1 000	750	550	750	400	400	450
Permissible speed on at least one section	200 km/h or more	Up to 200 km/h	Up to 160 km/h	Up to 140 km/h			
Planning quality	105 %	108 %	110 %	120 %	120 %	120 %	108 %
Category of train	Intercity express (ICE)	EuroCity and InterCity	InterRegio, main-line express trains	Express night trains, accompanied-car trains, fair trains	Regional express train, through train	Regional train, CityBahn, slow train	S-Bahn train

Table 2. **Percentage reduction in terms of train-km per year**

Percent reduction	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Main-line traffic train-km (millions)	14	28	42	56	70	84	98	112	126	140	154	168	182	196	210	224	238	252	266	280
Short-haul traffic train-km (millions)	0.3	3	25	63	134	205	250	293	333	370	407	444	481	518	555	592	629	666	703	740

Customers who order slots for several years and sign a contract are granted a further price reduction in addition to that mentioned above. For firm orders extending over 2, 3, 4 and 5 years, the corresponding reductions are 2, 3, 4 and 6 per cent.

The German pricing policy makes the financing of rail activities truly transparent, even though the determination of the costs in terms of train-km is far from receiving unanimous approval and constitutes an obstacle to the entry of new operators. The relatively high prices and the choice of pricing applied to the train rather than the wagon are dissuading new operators from moving in.

These price scales have introduced a certain flexibility, but it is still insufficient and, in a way, is institutionalising the status quo by discouraging the adoption of new techniques and limiting the freedom of manoeuvre of possible new entrants. Separation has not progressed very far because DB AG is still both service provider and network operator. The transition is a gentle one. The undertaking seems to have been genuinely successful since already more than sixty transport operators have moved onto the DB AG's rail network and their number is steadily increasing

While proposing rather high marginal network entry costs, this pricing system reduces, on the one hand, the uncertainty for future operators by encouraging longer-term commitments and a search for financial and technical partners and, on the other hand, short-term practices which could have pernicious effects on the continuity and quality of the rail service.

Finally, the choice of a high level of global pricing guarantees the infrastructure manager a development capability, which is one of the strong political choices of these reforms, together with the choice of a user tariff likely to lead to the optimal utilisation of the network.

4.3. The French example: a transitional phase[32]

Introduced against a background of social strife, in particular a big strike in 1995, the French reforms consisted in establishing *Réseau Ferré de France* (RFF), a new public company which, as balance-sheet liabilities, received three-quarters of the debt of the SNCF and, on the asset side, the national network infrastructure, with the exception of the stations and installations needed by the historical operator. The latter was entrusted with the management and maintenance of the infrastructure on behalf of the RFF, which pays the bill for this service (16.8 billion francs for 1997, the first year of implementation).

The SNCF pays the RFF for the use of the infrastructure. For the first two years, a limit was placed on this fee (slightly under 6 billion in 1997) by the law and the decree establishing the new system.

The first characteristic of the system relates to the fact that, relieved of most of its debt and infrastructure financing, the SNCF is in a position to balance its accounts, which it is expected to do in 1999. Secondly, the new infrastructure company, which at present can only count on earning 6 billion francs, must cover, in addition to nearly 17 billion in network maintenance and management costs and unavoidable investment costs amounting to about 13.6 billion, charges of around 9 billion on the debt inherited from the reforms. Obviously, most of the difference between expenditure and income is covered by the Government, in the form of either a capital grant or subsidies, the difference being made up by borrowing.

This, then, is a very special situation which can be interpreted in two different ways. Either the RFF may be regarded as a body whose principal function is to take over the debt and cover the deficit (net of subsidies) of the infrastructure account by borrowing. Naturally, in this case, investment would be the adjustment variable and would inevitably face historical decline. Or the present situation may be regarded as a transitional phase for putting new structures in place, after which pricing that offers greater incentives and ensures better coverage of the costs will be gradually introduced.

Being capped in 1997 and 1998, the present fee system is obviously far removed from the principle of covering the costs. It corresponds to about one-fifth (in terms of the total amount) of the German system. There is little connection between the six billion constraint imposed and the reality of the actual costs, particularly as more than half of this sum comes from the regional organising authorities (mainly “Parisian” passenger transport) and less than half from the SNCF.

There is little point in studying this provisional system, precisely because it is capped. However, it should be noted that the idea was to create incentives, especially where the demand for slots is high relative to capacity, i.e. on the urban and suburban lines (the part of the network designated R_0) and to a lesser extent on the busy high-speed lines (R_1). On the other hand, on the low-density, high-speed network and on the main-line network (R_2) the fees are very low, while on the rest of the network (R_3) they are symbolic.

The fee system distinguishes between a monthly access charge, AC, per kilometre of lines for which access is requested, a reservation charge, RC, per kilometre and per slot reserved and a traffic charge, TC, per train-kilometre.

There are different reservation charges for peak periods, normal periods and slack periods. The corresponding charges for 1997 are shown in Table 3.

Clearly, this system is much less detailed and sophisticated than that introduced on the German network and thus raises the question of whether it is sufficiently refined to enable the relevant marginal costs and homogeneous demand segments to be distinguished in the event of the future system being steered towards a more determinedly economic form of pricing.

Table 3. **French provisional pricing**
(1997 unit prices in francs)

Sub-network	$\mathbf{R_0}$	$\mathbf{R_1}$	$\mathbf{R_2}$	$\mathbf{R_3}$
AC	11 000	11 000	250	0
RC (peak)	100	18	0.85	0
RC (normal)	44	6	0.85	0
RC (slack)	20	4	0.85	0
TC	0.30	0.30	0.30	0.30

Note: AC is expressed per month and per km of line, RC per slot-km, TC per train-km.

In the French case, clearly the main problem is how pricing will evolve after 1999. This question is overshadowed by the fact that, overall, the rail system is running at a loss. The first step then will be to choose between two strategic directions. One choice would be a low-toll system which would concentrate the public contribution on covering the deficits of the RFF and financing new investment. In this case a policy of long-term marginal cost pricing without budgetary constraint might be envisaged. A second choice would be a system that combined budgetary constraint with a Ramsey-Boiteux principle. In this case the SNCF would have to be subsidised for a fairly long time, but the subsidies could be correlated with the loss-making services thus financed, thereby allowing the authorities latitude to compare their cost and their utility.

On the basis of a study in progress, the RFF is to propose to the Government a user pricing system designed to encourage a better allocation of resources.

5. CONCLUSION

To reach a conclusion on such a subject would be to suppose that a definitive theoretical contribution, which was both coherent and pertinent and proposed measurable concepts, would make it possible to solve, down to a few details, this difficult rail infrastructure pricing problem. Only a patient approach that takes into account all the attempts to apply theoretical prescriptions will enable us to work towards a satisfactory solution.

It is no insult to the achievements of economic theory or railway economics to conclude with the following few lines which were written about a century and a half ago:

We merely wished to show that the way in which the tolls are fixed can greatly extend the utility of certain routes and that the guiding principle in assessing these charges should not be to set a price proportional to the weight or the distance nor to favour a particular industry or a particular class of passengers, but rather to impose on each passenger and on each good only a price that is lower than that which would prevent the passenger or good from using the route. Admittedly, the methodical classification of these passengers and goods does call for inventiveness and an intimate knowledge of the local circumstances, but a sound theory can do much to facilitate this work[33].

NOTES

1. The pricing of the various transport sector modes should also take into account the environmental social costs and propose coherent pricing rules. This presentation does not expand on this point, which would require special development. We note that in this respect it is necessary to distinguish between two independent types of questions. On the one hand, rail pricing should be considered in relation to the social costs actually taken into account in the pricing of the principal competing mode, namely, road transport. This concern for coherence might even lead to the legitimising of intermodal balancing subsidies, theoretically justified by a second-order optimum. On the other hand, and this is a totally different problem, it is necessary to introduce differential pricing within the rail mode in order to take into account the advantages and disadvantages of the different technologies used by the operators and gradually encourage the use of those that are less polluting.

2. Brooks, M. and Button, K. propose a detailed typology of these various experiences in "Separating transport track from operations: a typology of international experiences", *International Journal of Transport Economics*, Vol. XXII, No. 3, October 1995.

3. This is the case in the United States for certain passenger services for which Amtrak must negotiate an infrastructure access charge with the integrated private operators.

4. The calculation methods used in these studies can often be reduced to very simple procedures, especially as the quality of data needed for more sophisticated calculations is very difficult to obtain. Nevertheless, the statistical analysis can be made more precise by greater refinement and by establishing precise relations between particular types of costs, networks and users.

5. Avoidable fixed costs are then defined as those which would disappear if the firm stopped producing one of its products.

6. Winston, in a survey of road pricing, explains that the proposal to consider optimal pricing and optimal investment in parallel harks back to the work of Herbert Mohring (1962).
This school includes such authors as T.E. Keeler, K.A. Small, M. Kraus, S. Glaister and S.A. Morrison.
Morrison, S.A. (1986), "A Survey of Road Pricing", *Transportation Research*, Vol. 20A, No. 2, March, pp. 85-97.
Mohring, H. and Harwitz, M.I. (1962), *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston, IL.
Small, Kenneth A., Clifford Winston and Carol A. Evans (1989), *Road Work*: *A new Highway Pricing and Investment Policy*, Washington, D.C., The Brookings Institution.
Hau, Timothy D. (1992), *Economic Fundamentals of Road Pricing: A Diagrammatic Analysis,* Transport Division, Infrastructure and Urban Development Department, The World Bank, December.

7. Allais, Maurice (1948), *Revue d'Économie Politique*, *op. cit*., p. 230.

8. Boiteux, Marcel (1994), "Le développement de l'approche économique du service public", *L'Europe à l'épreuve de l'intérêt général*, Collection ISUPE, Édition ASPE EUROPE, Paris, p. 46.

9. Kolm, Serge-Christophe (1968), "La théorie économique générale de l'encombrement", *Futuribles,* SEDEIS.
See also the work of H. Lévy-Lambert, notably: Levy-Lambert, Hubert (1968), "Tarification de services à qualité variable -- application aux péages de circulation", *Econometrica*, July-October, Vol. 36, No. 3-4, pp. 564-574.

10. Although traffic management is generally based on rules of priority, which distinguishes the rail problem from that of the roads and creates a further difficulty for the theoretical approach to the pricing of rail infrastructure.

11. Here we have used the demonstration proposed by Steven A. Morrison, who bases himself on the work of authors such as Mohring, Harwitz and Vickrey.

12. Hotelling, H. (1938), "The General Welfare in relation to problems of Taxation and of Railways and Utility Rates", *Econometrica*, 6(3), July, pp. 242-269.

13. Kolm, Serge-Christophe (1971), *La valeur publique, l'État et le système des prix,* CNRS, Dunod, Paris, p. 7.

14. In the literature, the French school is associated with the rejection of pricing based solely on balancing the accounts. This takes little account of the analyses of Jules Dupuit in the last century or of the more recent work of Maurice Allais and Marcel Boiteux.

15. Blaug, Mark (1986), "La pensée économique, origine et développement", *Economica*, 4th Edition, Paris, p. 718.

16. Allais, Maurice (1948), "Le problème de la coordination des transports et la théorie économique", *Revue d'Économie politique,* 58, pp. 212-271.

17. Vickrey, W. (1948), "Some Objections to Marginal Cost Pricing", in *Journal of Political Economy*, 56, pp. 218-238. (Reedited by Richard Arnott, Kenneth Arrow, Anthony B. Atkinson and Jacques H. Dreze, William Vickrey, *Public Economics,* p. 217).
This is also the view of the Forecasts Directorate: *"In situations in which increasing returns are found to exist, there will be a conflict of objectives between the perfect orientation of consumer demand and the proper accountability of the managers. In no circumstances should this second factor be neglected." Note Prévision*, Commissariat Général du Plan, p. 151.

18. Bernard, A. (1983), "Coût marginal ou coût moyen ?", in Quinet, E. (editor), *Les transports et la puissance publique*, Presses de l'École Nationale des Ponts et Chaussées, report on transport economics seminar organised in 1981-82 by the École Nationale des Ponts et Chaussées, p. 36.

19. Very many commentaries on this type of pricing point out that fairness implies that the consumers should bear the costs of producing the goods they consume and that all the consumers should pay the same unit price for the same good. This is a frequently recurring complaint: "those systems which differentiate between deficit tolls according to

the characteristics of the demand are generally considered unreasonable and unfair." Oort, C.J., *op. cit.*, p. 62.
Clearly, the theoretical considerations advanced here shatter this principle. The tolls applied may vary for products that are identical both from the technical standpoint and by reason of their cost.

20. Dupuit, Jules (1849), "De l'influence des péages sur l'utilité des voies de communication", *Annales des ponts et chaussées*, p. 248.

21. Boiteux, Marcel, "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire", *Econometrica*, No. 24, 1956, pp. 22-40.

22. Ramsey, F. (1927), "A contribution to the theory of taxation", *Economic Journal*, Vol. 37, No. 1.

23. Boiteux, M., *op. cit.*, p. 35.

24. With that "s" Maurice Allais distances himself from the standard model. For further details, see: Allais, Maurice (1971), "Les théories de l'équilibre économique général et de l'efficacité maximale. Impasses récentes et nouvelles perspectives", *Revue d'Économie Politique*, No. 3, May-June, pp. 331-409.

25. Dupuit, Jules (1849), "De l'influence des péages sur l'utilité des voies de communication", *Annales des ponts et chaussées*, pp. 170-248.

26. We shall not describe the reforms themselves as they have already been extensively analysed. See, for example:
    - Montagu (1995), "L'accès à l'infrastructure ferroviaire, l'expérience britannique", *Rail International*, 26, January, pp. 5-15. Paper read at the AICCF/CCFE Conference, Berlin, 6-7 October 1994: The Railways and European Transport Policy.
    - (1994), *Britain's Railways: a new Era*, The Department of Transport, London, March.
    - Preston, John (1996), "The economics of British Rail Privatisation: an assessment", *Transport Review*, Vol. 16, No. 1, pp. 1-21.
    - British Ministry of Transport (1996), *Transport Report*. Available on the Web at: <URL http://www.open.gov.uk/dot/ann_rpt/chap-1.htm>.
    - Nash, Christopher (1996), "The British Experience", *The Separation of Operations from Infrastructure in the Provision of Railway Services*, Round Table 103, ECMT, pp. 59-102.

- -- Bradshaw, B. and L. Mason (1994), *Rail Privatisation: Facts, Issues and Opportunities,* Oxford Economic Research Associates Ltd., Oxford.
- -- Baumstark, Luc, Laurent Guihéry, Florence Lacaille (1997), "Process of Deregulation within the Railways Sector: Access Pricing and Institutional Organisation", World Conference Research Rail 97, Florence.

27. Railtrack derives its income from user fees paid by the operators (supply of electricity, etc.), rents paid for the use of stations and depots and rents from its commercial assets. To these should be added the access charges which are determined by negotiation (see below). The procedures have been progressively refined. At the beginning, no rules for calculating the charges were laid down. The first charges were fixed at a level that would cover the total costs and ensure a return on the invested capital of the order of 8 per cent.

28. The charge includes a fixed annual fee comprising the allocated fixed costs (joint costs) and the additional fixed costs (specific to each company). The fixed charges, which correspond to about three-quarters of the infrastructure costs, form the subject of negotiations between the operators and Railtrack. The variable charges contain infrastructure user fees calculated in terms of train-miles which are different for each category of rolling stock (10 per cent of total costs). The costs incurred at regional and national levels are shared out among all the franchise holders in proportion to their receipts from fares. The costs incurred at local level, or on a single line, must be distributed among the users in proportion to the number of vehicle-kilometres travelled.

29. Avoidable costs: rule for the allocation of the fixed costs of the whole of the services provided by an operator, equal to the amount saved in the event of his services being eliminated.

30. Additional costs: increase in infrastructure costs imposed by its services, taking into account the configuration of the other services.

31. The corresponding subsidies are financed from revenue generated by the petroleum tax *(Mineralölsteuer).* Note that article 4 of the Railways Restructuring Act states that, from 1996, the DB AG will no longer receive any direct funds from the Federal Government for managing regional passenger services. The subsidies are allocated to

the *Länder*, which use them in accordance with their own regional transport policy. However, the *Länder* must use these transfers for public transport purposes.

32. See the following official texts:
    - -- Law No. 97-135 of 13 February 1997, establishing the public corporation, *Réseau Ferré de France*, with a view to the revival of rail transport;
    - -- Decree No. 97-446 of 5 May 1997 on national rail network user fees;
    - -- and, finally, the Orders of 30 December 1997 on national rail network user fees, JO, 31 December 1997, pp. 19461-19463.

33. Dupuit, Jules (1873), “Poids et mesure”, in Coquelin and Guillaumin (eds.), *Dictionnaire de l’économie politique*, 4th Edition, Vol. 2, Paris, pp. 339-344.

BIBLIOGRAPHY

Allais, Maurice (1945), *Économie pure et rendement social*, Edition Sirey, Paris, 72 pp.

Allais, Maurice (1948), “Le problème de la coordination des transports et la théorie économique”, *Revue d’Économie politique*, 58, pp. 212-271.

Allais, Maurice *et al.* (1965), Options de la politique tarifaire dans les transports, Report prepared at the request of Commission of the European Economic Community, *Studies Coll., Transport Series*, *No. 1*, Brussels.

Allais, Maurice (1989), “L’économie des infrastructures de transport et les fondements du calcul économique”, *Revue d’Économie Politique*, 99, No. 2, pp. 159-197.

Allais, Maurice (1989), *Théorie générale des surplus*, PUG, Presse Universitaire de Grenoble (2nd edition), 716 pp.

Banister, David, Kenneth Button *et al.* (1993), *Transport, the Environment and Sustainable Development,* Alexandrine Press, Oxford, 275 pp.

Baumstark, L. (1997), “Tarification de l’usage des infrastructures et théorie de l’allocation optimale des ressources, de la logique des coûts à la révélation des préférences”, Doctoral thesis, Université Lumière Lyons 2.

Boiteux, Marcel (1956), “La vente au coût marginal”, *Revue Française de l’Énergie*, December.

Boiteux, Marcel (1956), “Sur la gestion des monopoles publics astreints à l'équilibre budgétaire”, *Econometrica*, No. 24, pp. 22-40.

Bowers, Ph. (1995), “Commercialisation, accès local et international aux réseaux ferrés”, *Public Transport International,* No. 1.

Brooks, M., K. Button (1995), “Separating transport track from operations: a typology of international experiences”, *International Journal of Transport Economics*, Vol. XXII, No. 3, October.

Council Directive 91/440 of 29 July 1991 on the development of the Community's railways, OJ L 237 of 24/08/1991, 25 pp.

Council Directive 95/18/EC of 19 June 1995 on the licensing of railway undertakings (OJ L 143/70-74).

Council Directive 95/19/EC of 19 June 1995 on the allocation of railway infrastructure capacity and the charging of infrastructure fees (OJ L 143/75-78), 27 June 1995.

Regulation (EEC) No. 1191/69 of the Council of 28 June 1969 (OJ L 156) on action by Member States concerning the obligations inherent in the concept of a public service in transport by rail, road and inland waterway, last amended by Regulation No. 1893/91 of the Council, OJ L 169 of 29/6/1991.

Regulation (EEC) 1192/69 of the Council of 26 June 1969 on common rules for the normalisation of the accounts of railway undertakings, OJ L 156 of 28/06/1969.

Regulation (EEC) No. 1107/70 of the Council of 4 June 1970 on the granting of aids for transport by rail, road and inland waterway, OJ L 130 of 15/06/1970.

Green Paper (1995), *Towards fair and efficient pricing in transport,* COM(95)691 of 20/12/1995, 58 pp.

White Paper (1996), *A strategy for revitalising the Community's railways*, COM(96) 421 final, Brussels, 30 July 1996, 54 pp.

ECMT (1995), *Why Do We Need Railways?,* Proceedings of International Seminar, Paris, 19-20 January, 226 pp.

ECMT (1993), *Privatisation of Railways*, *Round Table 90*, Paris.

ECMT (1997), *The Separation of Operations from Infrastructure in the Provision of Railway Services*, *Round Table 103*, Paris.

Commissariat Général du Plan (1995), *Transport, le prix d’une stratégie* (Workshop chaired by A. Bonnafous), Vols. 1 and 2, La Documentation Française, Paris.

Commissariat Général du Plan (1996), *Transformation des opérateurs publics de réseaux en Allemagne -- Les cas du contrôle aérien, des transports ferroviaires, de la poste et des télécommunications,* Study made by Cabinet ESL & Network, Vols. 1 and 2, March.

Commissariat Général du Plan (1993), *Transports : pour une cohérence stratégique*, Workshop on the strategic directions of transport policy and their medium-term implications, Commissariat Général du Plan, September.

Commissariat Général du Plan (1994), *Transports : pour un meilleur choix des investissements* (Group chaired by Marcel Boiteux), La Documentation Française, November, 132 pp.

Commissariat Général du Plan (1995), *Transports : le prix d'une stratégie* (Workshop chaired by Alain Bonnafous), Vols. 1 and 2, La Documentation Française, pp. 213 and 182.

DB AG (1996), "Der Trassenpreiskatalog der DB AG: ein Vorbild für Europa?" Lecture given by Mrs Haase at the 29th Scientific Transport Research Seminar, "Competition Policy in Deregulated Transport Markets", organised by the University of Freiburg (Germany), Hinterzarten, October.

DB AG, *Trassenpreise -- Güterverkehr*, Frankfurt am Main, 1995.

DB AG, *Trassenpreise -- Personenverkehr*, Frankfurt am Main, 1995.

Dodgson, J. (1994), "Railway privatisation", in: M. Bishop, J. Kay and C. Mayer (eds.), *Privatisation and Economic Performance*, Oxford University Press, Oxford.

Drèze, H. (1994), "Marginal-Cost Pricing", Introduction to Part III in: Arnott, Richard, Kenneth Arrow, Anthony B. Atkinson and Jacques Dreze (eds.), *Public Economics* (Selected papers by William Vickrey), Cambridge University Press, pp. 190-195.

Ewers, H.-J. (1993), Privatisierung und Deregulierung bei den Eisenbahnen: das Beispiel der DB und der DR, Institut für Verkehrswissenschaft, Universität Münster, 15-17 September, 11 pp.

Gillen, David W. and Tae Hoon Oum (1991), *Transportation infrastructure: Pricing, Investment and Cost Recovery,* Royal Commission on National

Passenger Transportation, Technical Seminar, Ottawa, Canada, September 26 and 27, 181 pp.

Guihéry, L. (1996), "Fédéralisme allemand et réforme des chemins de fer en Allemagne : le transport ferroviaire régional de voyageurs devient une nouvelle compétence des Länder", *Documents-revue des questions allemandes*, No. 1, Paris, p. 79.

Guihéry, L. (1997), "Réforme des chemins de fer et tarification d'usage des infrastructures ferroviaires en Allemagne, une organisation contrôlée de la concurrence sur le réseau ferré allemand", *Transport,* No. 382, March-April, pp. 112-121.

Hau, Timothy D. (1993), "Economic Fundamentals of Road Pricing: A Diagrammatic Analysis", Working Paper, WPS 1070, Transport Division, Infrastructure and Urban Development Department, The World Bank, p. 87.

Horton, Sir Robert (1997), "Re-engineering Britain's railways: risks and rewards", *European Railway Review*, June, pp. 89-95.

Hotelling, H. (1938), "The general welfare in relation to problems of taxation and of railways and utility rates", *Econometrica*, 6(3), July, pp. 242-69.

Hutter, R. (1950), "La théorie économique et la gestion commerciale du chemin de fer", *Revue Générale des Chemins de Fer*, February, July and October, p. 44.

Kolm, Serge-Christophe (1968), "La théorie économique générale de l'encombrement", *Futuribles*, SEDEIS, p. 82.

Laffont, Jean-Jacques (1977), *Effets externes et théorie économique*, Monographies du séminaire d'économétrie, No. 13, éditions du CNRS, Paris, 200 pp.

Laffont, Jean-Jacques and Jean Tirole (1994), *A Theory of Incentives in Procurement and Regulation,* The MIT Press, Cambridge, Massachusetts, London, 705 pp.

Le Duc, M. (1995), *Services publics de réseau et Europe : Les exemples de l'Allemagne, de l'Espagne, de l'Italie et du Royaume-Uni*, Les études de La Documentation Française, Paris, 1995.

Morrison, Steven, *The Theory of Optimal Highway Pricing and Investment: Comment*, Northeastern University, Boston, Massachusetts.

Nash, C. (1993), “Developments in transport policy: rail privatisation in Britain”, *Journal of Transport Economics and Policy*, Vol. XXVII, No. 3, September, pp. 317-322.

Office of the Rail Regulator, *Framework for the approval of Railtrack's track access charges for freight services*, a consultation document, October 1994, 19 pp.

Preston, J. (1996), “The economics of British Rail privatisation: an assessment”, in *Transport Reviews*, Vol. 16, No. 1, pp. 1-21.

Quinet, E. (1990), *Analyse économique des transports*, PUF, Collection Économie, Paris, 302 pp.

Allais, Maurice, Del Viscovo, Duquesne de la Vinelle, Oort and Seidenfus (1965), “Options de la politique tarifaire dans les transports”. Report prepared at the request of the Commission of the European Economic Community, *Studies Collection, Transport Series*, No. 1, Brussels.

Rathery, A. (1984), “De nouvelles perspectives pour la politique des infrastructures de transport dans les Communautés européennes ? Possibilités offertes par une approche pragmatique de la tarification d'usage des infrastructures”, *Transports*, No. 297, October, pp. 358-365.

Rothengatter, W. (1991), “Deregulating the European Railway Industry: theoretical background and practical consequences”, *Transportation Research*, Vol. 25A, No. 4, pp. 181-191.

Small, Kenneth A. (1990), “Highway Infrastructure: Crisis in Finance or Crisis in Management”, *Review of the University of California Institute of Transportation Studies*, Vol. 13, No. 2, February, pp. 4-8.

Small, Kenneth A., Clifford Winston and Carol A. Evans (1989), *Road Work: A New Highway Pricing and Investment Policy*, The Brookings Institution, Washington, D.C.

The Department of Transport (1993), *Rail Freight Privatisation (the Government's proposals),* London, 27 pp.

Vickrey, W. (1987), “Marginal- and Average-Cost Pricing”, in: Eatwell *et al.* (eds.), *The New Palgrave Dictionary*, Vol. III, Macmillan, pp. 311-318.

Vickrey, W. (1994), *Public Economics* (Selected papers by William Vickrey), Edited by Richard Arnott, Kenneth Arrow, Anthony B. Atkinson and Jacques Dreze, Cambridge University Press, 558 pp.

Walters, A.A. (1987), “Congestion”, *The New Palgrave Dictionary*, pp. 570-573.

Wilkinson, A.W. (1992), “La situation politique des British Railways”, *Le Rail*, July, No. 33, p. 28-2.

UNITED KINGDOM

John DODGSON
National Economic Research Associates
London
United Kingdom

Acknowledgement

I am grateful to my colleague at NERA, Stuart Holder,
for valuable comments on an earlier draft.
However, I alone am responsible for the contents of this paper.

SUMMARY

London, November 1997

1. INTRODUCTION

This paper is concerned with the issue of user charges for railway infrastructure. The approach of the paper is to consider the principles that might underlie a set of user charges, rather than to provide detailed descriptions of actual systems of user charges. However, reference will be made to the British approach in a number of cases to show the methods which have actually been used in practice in applying principles of rail access charging.

There are a number of major questions which are relevant when considering rail access charges. These include:

- What do we mean by rail infrastructure, and what is the nature of the costs involved in providing it?
- What is the purpose of user charges for rail infrastructure services? What objectives are they intended to achieve in practice?
- Given the objectives for rail infrastructure charges, how can they be designed in practice to achieve these objectives?
- To what extent should user charges reflect costs?
- What theoretical concepts of costs are relevant: short-run marginal costs, long-run marginal costs, short-run average costs or long-run average costs?
- How can these costs be measured in practice on real rail systems?
- Who should pay the charges?
- Who should pay for those elements of rail infrastructure costs which are not recovered by means of charges?
- What role will administrative allocation mechanisms continue to play in a charge-based system?

In considering rail infrastructure charges, it should be noted that there are a number of mechanisms by which charges might be levied. These include:

- Standard published tariffs which are paid by all users of particular types of infrastructure services;

- Negotiated tariffs, agreed between train operators and infrastructure providers (these tariffs will usually not be revealed publicly);
- Systems whereby scarce access rights are auctioned in some way to the highest bidder or bidders. Auction systems are complex, though they can in principle provide a mechanism to extract the maximum value from a given set of rail infrastructure assets, and to point the way to the most effective improvements in those assets by means of investment.

Charges are also likely to continue to be subject to external review or revision by government or regulatory bodies. One issue likely to be of concern is the issue of avoiding charges which discriminate between train operators performing similar tasks. In addition, both the structure and levels of charges are likely to be subject to review. In Great Britain, the independent Rail Regulator, supported by his officials in the Office of the Rail Regulator (ORR), regulates both the structure and levels of access charges, as well as the access agreements themselves[1]. The level of track access charges for the main passenger services is regulated through a rule which has been applied across a number of formerly publicly-owned industries in Britain, namely, telecoms, gas, electricity, water and airports. This rule is the (RPI-X) rule, in which regulated prices are allowed to rise by no more than the Retail Price Index, RPI, *less* an element X which allows for the scope by which the Regulator believes the industry has to reduce its unit costs in real terms each year. The value of X is set at the beginning of each review period, which normally lasts five years, and is then normally held constant over the review period[2]. Towards the end of this period, there is a further review, when X is set for the following five years. The aim of this rule is to prevent the regulated firm from making excess profits, while at the same time preserving incentives, in that if the regulated firm cuts costs below X per cent per annum, it can keep the gains until the next review.

2. WHAT IS RAILWAY INFRASTRUCTURE?

In this chapter, we consider what railway infrastructure consists of. This is necessary in order to be able to move on in later chapters to consider the nature of the costs of its provision, and how these costs will vary with traffic levels both in the short- and the long-term, but also because there may be some differences of definition between different railway systems across Europe, and such differences might create confusion in discussion unless such differences are made explicit.

By infrastructure, we mean the infrastructure on which trains run, which we can classify under the following headings:

1. *The line of route.* This takes the form of the land and track bed itself, including necessary earthworks in the form of embankments and cuttings, plus civil engineering structures such as bridges and tunnels. Although some major structures may need renewal, much of the capital expenditure in this category will be sunk, and cannot be recovered. However, this does not mean that there will not be significant maintenance costs on an on-going basis. The earthworks will need to be kept in good order, particularly through maintenance of drainage facilities, while vegetation will need to be tended. Steel structures will need to be maintained by painting, while brickwork will need some attention as time goes by.

2. *The track.* The track itself consists of the ballast, sleepers, fastenings, rails, pointwork and crossings. They will need continuous inspection to maintain alignments, and in turn this will require tamping and other remedial action on the ballast. Rails and sleepers will need renewal from time to time, especially on heavily-used sections of route. There will be relatively low day-to-day operating costs, though wear-and-tear will be closely related to the actual numbers and weights of trains.

3. *Signalling.* The signal control system in operation on any section of line can vary considerably in complexity, from a single train in operation, to complex power signalling which, on high-speed lines, may involve some form of cab signalling system. The capital equipment in a conventional system will consist of fixed signals, track circuits, point control equipment, signal cabling or wires and signal boxes or control centres. Major renewal of equipment is likely to be infrequent, so that any network is likely to have very different vintages of technology in operation across the system. However, equipment is likely to require regular maintenance. In addition, day-to-day operation costs will reflect the labour requirements of the technology in operation and the complexity of the operation.

4. *Electricity power supply.* Where services are electrified, the electric supply equipment should also be regarded as part of the infrastructure. This equipment includes the overhead posts and catenary, or the third rail, the sub-stations and power supply cables, control equipment and (where provided by the rail infrastructure authority) any generating

equipment. A major element of costs will be that of the electricity itself, which might be bought off the rail infrastructure authority, or directly from the generator.

5. *Stations and depots.* The final element of the infrastructure is the stations, freight depots and rolling stock depots. These might be owned by the infrastructure authority itself, or be owned (or leased) by train operators. This paper is not primarily concerned with this particular aspect of rail infrastructure.

The five various components of rail infrastructure are shown in Table 1. As well as providing a brief description of the main assets under each heading, this table also shows the main characteristics of capital, renewal, maintenance and operating costs under each type of infrastructure cost.

Infrastructure operation also includes other cost elements which are not primarily associated with physical assets. These include the important functions of train planning and timetable production, as well as general management activity.

Table 1. **Rail Infrastructure Assets and Costs**

Type of asset	Description of assets	Capital expenditure	Major renewals	Maintenance costs	Operating costs
1. Line of route	Land, embankments, cuttings, bridges, tunnels	Original route construction	Major structure replacement	Culverts, vegetation, brickwork, steel structures	
2. The track	Ballast, sleepers, rails, switches and crossings	Ballast, track and pointwork	Renewal of sleepers, rails and other trackwork	Inspection, ballast tamping, partial repair of track	
3. Signalling	Fixed signals, track circuits, signal boxes, control equipment	Construction of signals, boxes, etc.	Replacement of signals, boxes, etc., usually with more modern equipment	Maintenance of signals and control equipment	Signal operations staff, plus power and other supplies
4. Electricity power supply	Overhead line or third rail, sub-stations, cabling, power control equipment	Electricity power supply and associated supply equipment	Replacement of supply equipment	Maintenance of supply equipment	Electric power, plus costs of control staff
5. Stations and depots	Platforms and buildings	Construction of stations and depots	Renewal or replacement of buildings	Maintenance of structures	Labour and other costs incurred by infrastructure authority or other organisation

3. WHY CHARGE FOR RAILWAY INFRASTRUCTURE?

Why should charges be levied for railway infrastructure? One simple answer is that charges are levied for most goods and services, and so should also be levied for railway infrastructure.

A more considered answer is that the market system -- through which charges are levied for goods and services throughout the economy to compensate the firms which provide them for the costs they incur in doing so -- ensures that society values the output of these goods and services more than it values the opportunity costs of the inputs used to produce them. That is to say, in the case of rail infrastructure, the inputs in terms of labour, energy, materials and capital used to provide and maintain railway track could alternatively be employed to produce other goods and services. By employing them in the railway industry, we give up these other goods and services. We therefore need to ensure that society values the provision of the rail infrastructure more highly than it values the goods or services which have been given up.

In the past, railway services were generally provided by single, vertically integrated firms. In such circumstances, there would be no explicit system of charges paid by the part of the firm which operated trains to the part which operated infrastructure. However, where the railway companies were required to operate commercially, in the days before government subsidy, railway customers would pay in charges sufficient in total to cover both the train operating costs and the infrastructure costs. Hence, the costs of rail infrastructure would be covered by rail users.

We next consider the reasons for having separate charges for rail infrastructure access.

A specific reason for levying charges for railway track capacity is to provide a mechanism by which different train operators can have access to the railway network on a comparable basis. This might enable new operators providing innovative services to improve the range and quality of transport services which railways can provide, and it may permit competition between train operators to reduce costs and improve service quality

However, where a rail system remains vertically integrated, the integrated rail company will be reluctant to permit new operators onto its network when they will compete with its own train services. It is clear that the new operator should make some contribution to the costs of infrastructure provision, but the infrastructure provider may try to charge a rate so high that it discourages entry or, at the least, puts the entrant at a comparative disadvantage[3]. In order to deal with this situation, American regulatory economists, William Baumol and Robert Willig, developed the Efficient Component Pricing Rule (the ECPR). Under this rule (see Baumol, 1983), the track access charge is set equal to the additional cost to the infrastructure provider of accommodating extra trains, plus the loss of profits from its own trains as a result of the increased competition.

However, the difficulties of securing a "level playing-field" for different train operators competing with an integrated supplier have increasingly led, mainly because of pressure from the European Commission under Directive 91/440, to a split between rail infrastructure and operations in Europe, both from an accounting and, increasingly, from an operational point of view. Once the infrastructure and operations parts of the railway company are separate, then the provision of infrastructure access to train operators becomes a market transaction.

We can distinguish four objectives of the resulting systems of charges. These are as follows:

1. *Efficient use of infrastructure.* Charges should ensure that whatever level of infrastructure is provided is used most effectively. This means that the value of the infrastructure to rail users should be maximised. A given amount of railway track might be used in a number of different ways, and an administrative allocation mechanism might not make the best use of that capacity. For example, train paths might be allocated to slow stopping trains, which prevent use of the capacity by more highly-valued fast trains which have greater earning potential. Or spare paths might be provided for occasional freights which rarely run, and which might be accommodated in a more effective manner if the freight operator had to pay the costs of the paths left available for this occasional use. Or there might be many alternative ways of allocating train paths through a complex network of junctions, and pricing systems might give better incentives to determine the best pattern of paths for the different types of train passing through the junctions than do administrative systems based on historic timetable patterns. Finally, pricing systems might provide a way to improve the quality of the output supplied by a particular block of infrastructure facilities.

2. *Efficient provision of infrastructure.* This means that the charging mechanism should lead to the provision of the correct level of rail infrastructure given the demands for rail use. Consequently, charges should give appropriate signals for investment in improvements in the infrastructure, and for contractions in the infrastructure to eliminate excess capacity (other than that whose retention is justified on the basis of anticipated increases in demand). In addition, the charging system should give incentives for the infrastructure provider to provide infrastructure of a specified quality at minimum cost: this means there should be no incentives for the infrastructure provider to employ excessive levels of labour or use excessive levels of capital.

3. *Efficiency in the rest of the railway business.* The infrastructure charges should also provide incentives for efficiency in the rest of the railway business. One important way they can do this is by providing a framework which encourages effective competition in the market for rail services, by encouraging competition where this will improve the range or quality of train services provided and/or reduce the costs of providing these services. Thus rail infrastructure charges can aid in the development of the most efficient market structure for both static and dynamic efficiency in the railway industry.

4. *Consistency with government objectives.* Such consistency can include consistency with fiscal objectives, such as an objective to reduce public subsidy for rail services, and consistency with transport policy objectives. The latter might include a desire to achieve efficiency in intermodal competition by means of user charges for both road and rail services, derived on an equivalent basis.

4. COST-BASED CHARGES

4.1. Concepts of costs

One of the most common principles of rail infrastructure charges is that they should be based on the costs of providing rail infrastructure. The *general* principle relevant here is that rail users should be prepared to pay the costs which they impose on society by diverting resources from other uses to the provision and use of rail infrastructure. However, there are a number of alternative cost concepts that can be relevant. These include:

- Short-run average costs: the total costs of providing infrastructure services at the existing level of capacity, divided by the number of traffic units[4];
- Long-run average costs: the total costs of providing infrastructure services, allowing for any required increase or decrease in capacity to produce services at lowest possible costs, divided by the number of traffic units;
- Short-run marginal costs: the extra cost of handling an extra unit of traffic on the existing system;
- Long-run marginal costs: the extra cost of handling an extra unit of traffic on the system when capacity can be changed so as to handle the increased traffic volume at least cost.

Although charges based on average costs would enable the rail operator to recover total costs, as long as demand were sufficient, optimum resource allocation will be achieved where charges are set equal to marginal costs. Where capital stock is optimally-adjusted, that is to say where the total cost of catering for the traffic on offer is minimised, short-run marginal costs will be equal to long-run marginal costs and so the marginal-cost-based pricing rule will have prices set equal to both SRMC and to LRMC. However, since traffic levels are always subject to change -- and these changes often cannot be forecast in advance -- and since railway infrastructure provision can only be changed slowly in response to traffic changes, the railway infrastructure in existence at any particular point in time may not be the level that is optimal for the traffic on offer. In these circumstances, prices set equal to short-run marginal costs will ensure the optimum use of the existing capacity. This is because capacity will only be allocated to those users who value the services that the infrastructure provides at least as much as the extra costs which need to be incurred to provide it.

However, even if short-run marginal costs can be measured, an issue which we will turn to in the next chapter, use of SRMC might lead to a set of charges which fluctuate over time. In contrast, LRMC-cost-based charges might be subject to less fluctuation and therefore can provide a more stable basis on which to plan future service and investment levels.

4.2. Measurement of costs in practice

In this chapter, we consider cost measurement in practice, with particular reference to the measurement of marginal infrastructure costs.

4.2.1. Short-run marginal costs (SRMC)

Short-run marginal costs show how the infrastructure costs change in the short-run when rail traffic levels change. In general, these changes tend to be low when traffic is not close to capacity limits. In such circumstances of uncongested rail networks, the main variations in infrastructure costs with traffic levels are the following:

- If the number and/or weight of trains increases, this can increase track wear-and-tear, and hence the maintenance expenditures necessary to keep track quality standards at a given level;[5]
- An increase in the number of trains might increase signal operation costs. For example, in an old-fashioned mechanical signalbox, the number of individual lever movements will increase as the number of trains increases. This will involve the signal person in extra effort, though it might not involve him or her in extra time since they have to be on duty for the period when the box is open. On the other hand, if signal boxes are not manned for 24 hours, extra trains might increase box opening hours. In addition, more complex operations might attract extra wage payments, and eventually more skilled staff may be required. In modern, computer-based signalling systems, additional movements may involve very little extra cost at all, once the system has been installed. (The complexity of the movements that have to be handled will determine the cost and complexity of the signalling system that is required, but this is an issue which relates to long-run marginal costs.);
- There will be administrative costs in running additional trains. In an extreme example, a "one-off" special passenger excursion may require considerable management effort in determining a path for the train through the system and in providing special notices to signalmen for the running of the train.

Where a rail system is near to capacity, short-run marginal costs reflect the opportunity costs of use of the capacity by the different trains on the system. There are a number of components of the opportunity costs of running a particular train. If a train runs in a particular path:

- it may displace another train to a less-preferred path. This path may be less preferred because it has less commercial potential (for example, it may disrupt a regular, clockface timetable, or it may offer a freight customer a less-desirable loading time) or it might reduce rolling-stock utilisation (for example, by increasing turn-round time at a terminal);

- it might increase journey times of other trains (for example, if they are delayed in crossing junctions);
- it may increase delays imposed on other trains if it runs late;
- it may mean that another train is displaced completely from the network.

The third of these types of congestion cost, namely, the delays imposed by an extra train on others, might be measured by means of a "performance regime", such as that operated in Britain by Railtrack. Under this system, Railtrack is required to pay penalty payments to train operators if their trains are delayed. On the other hand, a train operator whose own trains cause delays must make a penalty payment to Railtrack. Thus, delays imposed by one train operator on another require a penalty payment, effectively from one operator to another, channelled through the infrastructure provider, Railtrack. On the other hand, good performance is rewarded with bonuses.

In addition, the system includes the contractors who maintain the infrastructure, so that if delays are caused by their actions, Railtrack can recoup 80 per cent of the cost from them -- but not 100 per cent, so that Railtrack still has an incentive to encourage the contractors to improve their performance.

Train running throughout the system is monitored, and delays longer than three minutes are attributed to the party responsible[6]. According to Brian Mellitt, Railtrack's Director of Engineering and Production (Mellitt, 1997), the average payment per minute of train delay across a range of services varies between £17 for regional passenger services to up to £130-170 per minute for some South-East commuter services. These payments are based on average numbers of passengers per train, time values for rail passengers and average fares and fare elasticity values.

4.2.2. Long-run marginal costs (LRMC) and long-run incremental costs (LRIC)

Long-run marginal costs show the extra costs of providing an extra unit of traffic when the level of infrastructure provided can be adjusted. This long-run marginal cost for an extra unit of traffic in the form of a single train can be equal to short-run marginal costs (i.e. marginal capacity costs can be zero) if there is already enough capacity to handle the traffic at minimum cost. Expansions in capacity usually occur in order to accommodate a number of additional trains, rather than a single additional train. In these circumstances, we can distinguish long-run incremental costs, where the LRICs are the additional capacity and other costs of handling an extra block of traffic. These LRICs can be expressed in unit terms as long-run average incremental costs (LRAICs), which are equal to the LRICs averaged over the extra units of traffic handled by the additional capacity.

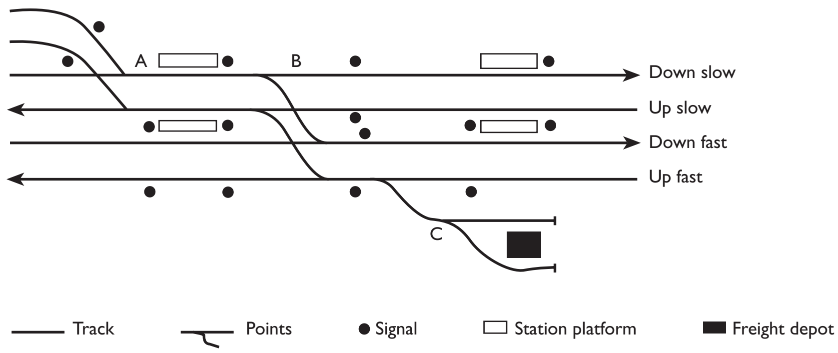
The use of LRICs in railway track costing can be illustrated by reference to their use in Britain. In Britain, LRICs have been used in two ways in railway infrastructure costing.

The first instance was in the allocation of total infrastructure costs between different train operators. Here, the need was to have a system of charges that covered the total costs of the commercial rail infrastructure company, Railtrack. An attempt was therefore made to allocate those costs which did not vary in the short-term with traffic levels, to the different train operators.

To do this, the network was divided into individual sections and the assets devoted to each section were identified. Unit cost figures were associated with each type of asset, such as length of plain line, 2-aspect signal, track circuit or point. The exercise then considered what assets would be saved in the long term if each operator's trains no longer used that part of the network: this long-term analysis considered reconfigurations of the route, such as that required to switch from 4-aspect to 3-aspect signals where traffic reductions warranted.

The approach is illustrated in Figure 1. This shows the track and signalling plan for a four-track route, used by four operators: a commuter train operator, an express service operator, an infrequent long-distance passenger operator and a freight operator.

Figure 1. **Calculation of long-run incremental costs for individual train operators**



The initial assessment of LRICs in the British railway industry asked the question:

> **What infrastructure costs would be saved if each operator's trains were withdrawn in turn?**

We consider these operators' characteristics and infrastructure needs as follows:

- The commuter operator operates a 15 minute headway service between the hours of 06.00 and 24.00 every day. This requires use of the two slow lines, though trains can be diverted to the fast lines in an emergency;
- The express service operator operates a 30-minute headway service over the fast lines between 07.00 and 23.00 every day;
- The infrequent long-distance passenger operator operates six trains a day in each direction. These join or leave the route at point A, and then cross from the slow lines to the fast lines at point B;
- The freight operator serves the freight depot at point C, using the fast lines in between the express operators' services.

We assume that the whole route section is signalled from a modern power signal box which controls a much longer section of interconnected routes.

The original LRIC approach would involve considering what assets would no longer be needed if each operator's trains were withdrawn. We next consider this in turn:

– If the commuter operator did not operate, the slow lines would not be needed at all, so the four-track layout could be reduced to a two-track one;
– If the express operator did not operate, the fast lines would not be needed, except for the link to the freight depot;
– If the occasional long-distance service did not operate, the junction and associated signalling at point A would not be needed;
– If the freight operator did not operate, the links and associated signalling for the freight depot would not be needed.

Each of these service withdrawals would be considered in turn and the cost implications calculated. Each of the resulting LRIC cost allocations was then allocated to one of the 24 mainland passenger Train Operating Companies or to another operator. When this exercise was carried out, about half the “fixed” track costs were allocated to LRICs and the rest had to be treated as common costs. However, this exercise was not then used to relate traffic levels to costs, since the block of LRICs from each section of route was then aggregated over all route sections to give a total LRIC allocation to each TOC. To this was added the allocation of common costs to that TOC, to give a total annual fixed track access charge to each TOC. TOCs cannot avoid these access charges by running fewer trains, so in effect they are simply a block sum which each TOC has to meet. The block sums differ from TOC to TOC, but because they were known in advance before the franchise bidding process, bidders factored them into their bid as a fixed cost that they would have to meet out of revenue. They have not been used to show how infrastructure costs vary from service-to-service as a way of seeing whether individual services are providing the taxpayer with value for money.

In part, this reflects the fact that the LRIC exercise described above was a “one-off” exercise for cost allocation. Only a limited number of service changes were considered, namely, those involving all of a particular operator’s trains from each route. In practice, costs might be saved by withdrawing or rescheduling only certain trains from a route, such as certain peak-hour trains, or trains at the beginning and end of the day, or certain trains that conflict with other trains that a new operator might wish to operate; such relatively small adjustments to services might lead to significant savings in costs at the margin.

Alternatively, substantial "blocks" of infrastructure costs might be saved if all trains were removed from a route. The overall social costs and benefits of this could be considered using some form of social cost-benefit analysis[7].

The second application of the LRIC approach in Britain arises where train operators are seeking additional train paths. Where this is the case, Railtrack will require that the train operator provides at least the extra costs of accommodating the additional train paths on the network, and where capacity is full, these costs will include additional capital costs. Now it is necessary to estimate long-run incremental costs on a much more incremental basis. In the British situation, the charges will normally be determined by a process of negotiation between the infrastructure provider and the train operator, though the Regulator can be called on to mediate.

4.3. The relationship between rail infrastructure costs and traffic levels: a statistical approach

One method to derive a closer relationship between rail infrastructure costs and traffic levels would be to use statistical analysis. In this approach, a cross-section of data on different route sections of a rail administration's system could be used to relate cost levels to traffic levels and to other factors, such as section length, which are believed to influence costs. This application of statistical cost analysis differs from the more common application, which is to compare total costs across different systems. The first requirement for such an approach would be appropriately-disaggregated cost and traffic data for route sections of a national network. Such data may already exist from some European systems: for example, Dodgson and Rodriguez Alvarez (1996) discuss cost allocation methods for RENFE data on infrastructure costs on individual route sections of the Spanish national railway system.

5. AUCTIONING CAPACITY

So far, we have considered systems of published tariffs and have mentioned a role for negotiated charges. One proposal for charging for railway track capacity we have yet to consider is the use of some form of auctioning mechanism. In this chapter, we outline the logic for using auction mechanisms and discuss the way they might be used for railway track capacity.

We first of all consider why auctions might be appropriate. Auctions are used, in practice, to sell a variety of products. A major reason for using an auction is that the seller may not have a very clear idea of what potential purchasers may be prepared to pay. An auction mechanism can then maximise the seller's return. This is true in the case of works of art, each of which is in some case unique, and in the case of perishable agricultural products and fish, where demand and supply are both subject to considerable variability. It is also relevant in the case of railway track, where markets are not yet developed, so that providers of railway track capacity may have very little idea of what different parts of the capacity are "worth" to different train operators. In addition, individual train paths are both unique and perishable.

Most auctions involve the sale of fairly clearly-defined lots, such as a particular painting, but a major problem with rail capacity is that a particular block of capacity often does not consist of a clearly-defined number of separate train paths, but can be parcelled up into different sets of non-conflicting paths, especially where train paths can conflict across flat junctions. Secondly, train operators will be interested in packages of slots, for example, to provide a regular-interval service of trains and/or to maximise utilisation of their rolling stock, so they will not be interested in bidding for individual slots. Third, their valuations of these packages may also depend on the services operated by competing and complementary operators. Fourth, the packages of slots allocated to the operators must be mutually compatible in that they avoid conflicting movements. Finally, depending on the form of the auction mechanism, bidders might have incentives to understate their true valuations of the track capacity they seek to occupy.

The problem of using auction mechanisms to allocate capacity in the transport sector has been considered in the case of airports. Here, airlines require packages of slots which will enable them to operate commercial services, but the availability of slots will be limited by runway (and terminal) capacity. Rassenti, Smith and Bulfin (1982) considered a system of combinatorial auctions in which airlines bid for alternative packages of slots on the understanding that they will only actually receive one of these packages. The airport then has to maximise its revenue from airlines while at the same time choosing a mutually-compatible set of bids from the airlines wishing to use its capacity. The mathematical models required to solve for the optimal bid are complex and, although there has been interest in the potential for auctions to solve this problem of allocating runway capacity, such systems have not been introduced in practice. One of the difficulties is that faced by airlines in deciding how to formulate the alternative bids that they might make[8].

Another problem in the airport case is that airlines need to negotiate with a number of infrastructure suppliers to build up a set of take-off and landing slots that enable them to operate their services between different airports. Except in the case of international services, train operators will be negotiating with a single national railway track supplier. However, in the railway case, potential conflicts between train operators are likely to be greater than in the airline case. This is because most airport take-off and landing slots are defined in terms of a 15-minute period, with the actual slot used within this period on any day being determined by air traffic control according to administrative/safety procedures. On the other hand, rail timetables, and certainly those for passenger trains, determine the precise ordering of train paths through a particular network, which makes allocation more complex.

One approach to the auctioning of railway track capacity is that developed by Lerz (1996). Lerz proposes the use of a system which is based on a second-price sealed bid auction. In this type of auction, developed by 1996 Nobel Prize winner in economics, William Vickrey, bidders provide sealed bids, but if they win, they only pay the price bid by the next highest bid. At first sight, this seems odd, in that the seller seems to be foregoing some revenue. However, it can be demonstrated that if bidders know they will only pay the second-highest bid, their optimal bidding strategy is then to reveal their true valuation of the good or service they seek to buy. This is because, if they bid more than their true valuation, they risk winning the auction *and* paying more than their true valuation: this would occur if the next highest bidder values the capacity at more than they do, so that they end up paying a second-price bid which exceeds their own true valuation. On the other hand, if they bid less than their true valuation, they may end up losing the auction in circumstances where they would have been prepared to pay a higher price than the winning bidder.

This sounds complicated enough, but in the case of railway track, the seller of railway track capacity will usually not be aware as to how best to parcel up the track capacity available into different "lots", since operators will usually not want to bid for individual train paths unless they can be sure of getting an acceptable package[9]. Lerz considers a pricing mechanism in which bidders pay a price which reflects the congestion costs imposed by each train operator on all other operators, including those priced off the system because of capacity constraints. These "congestion charges" are in addition to user fees which cover the infrastructure suppliers' costs in providing the capacity. They represent an application of a second-price auction, in that they are equal to the difference between the valuation which all other operators place on the use of the capacity in the absence of the operator paying the congestion fee, and the valuation which all other operators place on the actual use of the capacity: this gives a definition of congestion costs in terms of the opportunity costs which each operator's actual use of the capacity imposes on all other operators[10].

This still leaves the problem of deciding how to parcel up the capacity, but this auction system can help to do this because each operator will bid not for individual slots but for a complete timetable of train paths which includes not only his own paths, but also those of all other operators on the system. Each operator would therefore bid for alternative timetables in the knowledge that only one timetable would be operated in practice. It can be shown that each operator has an incentive to reveal his own valuations correctly, though this does not mean that operators would be convinced that they did indeed have such incentives. Implementing this type of system of auction for railway capacity in practice would be extremely difficult.

6. HOW SHOULD TOTAL COSTS BE COVERED?

A charging system based on average costs would recover total costs. However, the problem with such charging systems in industries like railways, where there are significant joint costs, is that some users may be prepared to pay a price which exceeds marginal costs, but not one which covers average costs. These customers will then be lost from the system even though they were prepared to pay economically efficient prices. Moreover, more of the joint costs will then be loaded onto the remaining users.

However, if charges are based on short-run marginal costs and there are no congestion charges, then track access charging revenue will fall well short of covering total costs. Even where charges were based on long-run marginal costs or long-run incremental costs, the British evidence suggests that revenue would not cover total rail infrastructure charges. The question which then arises is that of how total costs will be covered. The following are the main possible methods:

1. *Government subsidy for rail infrastructure*. The European Commission's Railways White Paper (Commission of the European Communities, 1996, p. 18) notes that the Commission's Green Paper on transport pricing proposes that, as a long-term objective, charges in the transport sector should cover both direct and external costs. While

confirming this as a long-term objective, the White Paper notes that in the short term it is legitimate for States to finance investment in railway infrastructure to compensate for unpaid external costs in the roads sector or to meet non-transport objectives. The danger with the use of government subsidies -- apart from the unanswered question of how long a period "the short term" is intended to cover -- is that care will be needed to ensure that the provision of subsidy does not blunt incentives on the rail infrastructure authority to secure efficiency in the provision of railway track and signalling. The infrastructure authority needs to provide infrastructure at minimum cost, and it also needs to provide the appropriate level of infrastructure for the traffic on offer. If the Government simply provides a block grant, equal to the infrastructure authority's deficit, then the authority may well be inefficient in its use of manpower, capital and other resources and it may carry excess capacity in rail infrastructure. The challenge is therefore to devise a subsidy system that provides the appropriate incentives.

2. *Multi-part tariffs*. With multi-part tariffs, train operators pay track access charges which reflect marginal costs, together with fixed sums which together will cover all the infrastructure operator's costs. These fixed costs would need to be designed so that they did not deter individual operators from providing any train services at all when these operators would be prepared to pay marginal infrastructure costs. Where train services cannot be justified commercially, but are justified on social grounds (for example, because of benefits in terms of reduced traffic congestion and/or reduced air pollution), then governments might pay, or contribute to, the fixed elements of the access fees. This is, in effect, what happens in Britain, where most of the passenger train operating companies receive franchise payments from the Government via its Office of Passenger Rail Franchising (OPRAF). However. one operator, Gatwick Express, receives no subsidy and so has to meet all the fixed -- as well as the variable -- elements of its track access charge itself, while other operators have contracted to go without subsidy by the end of their franchise period. The infrastructure authority, Railtrack, has incentives to operate efficiently because it is a commercial organisation whose profitability depends on its ability to control costs.
3. *Ramsey pricing*. This is a form of pricing advocated by public enterprise economists. Ramsey prices involve mark-ups above marginal costs, where prices exceed marginal costs so as to recover full costs from customers with minimum losses of economic

efficiency. To do this, the mark-ups are inversely related to the purchasers' elasticities of demand for the service, in this case rail track capacity, being sold. This is broadly equivalent to the old railway policy of "charging what the traffic will bear". Applied to railway track and signalling capacity, it involves charging a higher mark-up to those train operators whose customers are more captive to rail, such as regular commuters, and a lower mark-up to those train operators whose customers are less captive to rail, such as intermodal freight shippers.

4. *Some combination of the above*. In practice, cost recovery might be achieved by some combination of the above methods. Thus charges could cover marginal costs with some additional mark-ups based on ability to pay, though not ones sufficient to cover total costs. The remaining costs might then be recovered by a combination of fixed block payments and carefully-designed government subsidies.

7. WHAT IS THE RESIDUAL ROLE OF ADMINISTRATIVE ALLOCATION MECHANISMS?

In this chapter of the paper, we show how, even in a system based on infrastructure charging methods, administrative mechanisms still have an important role to play in allocating track capacity between different train operators. In effect, this is because no practicable track charging system could fully "clear the market", so that the timetable would be fully determined by the charging system.

To illustrate the use of administrative systems, we consider the mechanisms in operation in Britain. Train operators need to know what they will pay for railway track capacity, but they also need to have some guarantees that they will be able to obtain the types of capacity they will need to run their trains. This is achieved by means of track access agreements. These agreements between Railtrack and each train operating company require approval by the Regulator and set out the conditions under which access is provided. They are also public documents, and so can be consulted by other train operators or interested third parties.

Although the agreements differ between operators, that between Railtrack and Great Western is typical. This train operator operates high-speed diesel trains over a network of routes from London Paddington to the South-West of Britain and to Wales. The access agreement specifies maximum numbers of peak morning slots into Paddington and evening peak hour slots out. Apart from on the single-track Cotswold route, individual train paths are not specified, though the operator has rights to numbers of paths, to point-to-point journey times, to stops at certain specified stations, to operate relief trains, to use routes for empty stock movement, to use diversionary routes where the normal lines are blocked and to use routes to move stock to heavy maintenance facilities.

Thus, the access agreements do not in general specify particular train paths. Instead, the rights are exercised through the timetabling process which is specified in every track access agreement. The timetabling process which has been in operation since 1994[11] is essentially a consultative one in which:

- Railtrack produces the Rules of the Plan and the Rules of the Route, which set out key parameters for timetable planning and details of engineering works proposed. The Rules of the Plan detail timing capabilities of the network such as running times for different categories of trains, minimum headways between trains over different sections of route and stopping times at stations. The Rules of the Route deal with the timing and location of engineering possessions and permitted speed and other restrictions, for planned maintenance work;
- There is an iterative "bidding" process (but not one that involves payments) in which train operators bid for train paths consistent with their access rights, and Railtrack allocates capacity on the basis of "public interest decision criteria"[12]. At present, there are usually two such iterations for Mondays-Saturdays, but one for Sundays. Operators are likely to wish to bid 42-46 weeks in advance of the introduction of a new timetable coming into operation;
- An appeals procedure is used to settle disputes;
- Railtrack produces and publishes the passenger timetable twice a year; and
- Spare capacity within the timetable is allocated through a separate "spot bidding" process.

Though systems are likely to vary from railway to railway, some form of administrative system for timetable planning and the allocation of capacity will be needed to complement the allocation of capacity on the basis of track user charges, where there are a number of train operators competing to operate on the same network.

8. CONCLUSIONS

This paper has considered principles in determining track user charges for railway services. The most important principle is that charges should be based on costs, and the second most important is that the appropriate concept of costs is marginal costs. Where capacity is constrained, some form of congestion-based scheme would be appropriate where different operators are competing to use the capacity and this, in turn, might imply differentiation between peak and off-peak charges. However, although auctioning mechanisms might be used to determine these congestion charges in principle, it will be difficult to implement such mechanisms in practice.

In the meantime, charges should be based at least on short-run marginal infrastructure provision costs. This will leave a high proportion of total costs unrecovered, so attempts should be made to consider whether other elements of the costs of infrastructure provision on particular routes could be properly apportioned to individual train operators on the basis of causality. One approach worth investigating to do this would be the use of statistical analyses of rail infrastructure cost data.

Finally, charging systems will need to be accompanied by administrative systems for drawing up the operating timetables which are necessary for the safe physical operation of any railway system.

NOTES

1. The nature of these agreements is discussed in Chapter 7.

2. The initial value of X for Railtrack was set at 2 per cent in the Rail Regulator's first review. In addition, the Regulator imposed an 8 per cent cut in Railtrack's charges for passenger Train Operating Companies between 1994-95 and 1995-96. For more details of the original determination of Railtrack's charges, see Dodgson (1997).

3. The vertically-integrated railway might also simply be obstructive and refuse access, or make it impossibly bureaucratic.

4. There may be difficulties in practice in defining a traffic unit, where rail capacity handles different types of train.

5. In Britain, the costs of track wear-and-tear have been measured by track engineers, and can be shown to be related to the speed and axle weight of trains. These cost relationships form the basis of a set of published variable track charges which are levied by Railtrack on different types of rail vehicle-locomotives, multiple units, coaches and wagons. However, these charges only account for 3 per cent of total track costs.

6. Delays which are not attributed are allocated an unexplained code, and are then attributed to Railtrack by default.

7. Social cost-benefit analysis is to be used to consider changes in service specifications for passenger train operations in Britain: see Office of Passenger Rail Franchising (1997).

8. However, there is some formal (and informal) slot trading between airlines at some airports, and there may be potential in the railway industry for train operators to exchange or adjust train paths by mutual agreement.

9. Acceptable in the sense of enabling the operator to operate a commercially viable service.

10. This is more complex than the road congestion case, where each vehicle of a similar size imposes a similar time cost on all other vehicles in the traffic flow.

11. The Rail Regulator issued a Consultation Document seeking the industry's views on this process in October 1997.

12. These criteria consist of the following, none of which necessarily has priority over another:

    1) Sharing the capacity and securing the development of the network for the carriage of passengers and goods in the most efficient and economical manner in the interests of all users of railway services having regard, in particular, to safety, the effect on the environment of the provision of railway services and the proper maintenance, improvement and enlargement of the network;

    2) Enabling a Train Operator to comply with any contract to which it is party (including any contracts with their customers and, in the case of a Train Operator who is a franchisee or franchise operator, including the franchise agreement to which it is a party), in each case to the extent that Railtrack is aware or has been informed of such contracts;

    3) Maintaining and improving levels of service reliability;

    4) Maintaining, renewing and carrying out other necessary work on or in relation to the network;

    5) Maintaining and improving connections between railway passenger services;

    6) Avoiding material deterioration of the service patterns of operators of trains (namely, the train departure and arrival frequencies, stopping patterns, intervals between departures and journey times) which those operators possess at the time of the application of these criteria;

    7) Ensuring that, where the demand of passengers to travel between two points is evenly spread over a given period, the overall pattern of rail services should be similarly spread over that period;

8) Enabling operators of trains to utilise their railway assets efficiently and avoiding having to increase the numbers of railway assets which the operators require to maintain their service patterns;

9) Facilitating new commercial opportunities, including promoting competition in final markets and ensuring reasonable access to the network by new operators of trains;

10) Avoiding, wherever practicable, frequent timetable changes, in particular for railway passenger services; and

11) Taking into account the commercial interests of Railtrack and existing and potential operators of trains in a manner compatible with the above.

BIBLIOGRAPHY

Baumol, W.J. (1983), "Some subtle pricing issues in railroad regulation", *International Journal of Transport Economics,* 341-355.

Bolt, C. (1995), "Setting track access charges for the railways". Paper presented to the Conference on New Developments in Access Pricing for Network Utilities, 24 May 1995, London Business School. (Bolt is Chief Economic Adviser, Office of the Rail Regulator.)

Commission of the European Communities (1996), *White Paper: A Strategy for Revitalising the Community's Railways,* COM(96) 421 Final.

Dodgson, J.S. (1997), "Railway privatisation and infrastructure", in: D.A. Hensher and J. King (eds.), *World Transport Research*, Elsevier.

Dodgson, J.S. and P. Rodriguez Alvarez (1996), "Rentabilidad de los diferentes servicios de RENFE", in: J.A. Herce and Ginés de Rus, *La Regulacion de los Transportes en España,* FEDEA/Editorial Civitas, Madrid, pp 313-388.

Lerz, S. (1996), *Congestion Theory and Railway Traffic,* Labyrint Publications, The Netherlands.

Mellitt, B. (1997), "The new rail performance regime", *Modern Railways,* July, 433-437.

Office of Passenger Rail Franchising (1997), *Appraisal of Support for Passenger Rail Services. Planning Criteria: an Interim Guide,* OPRAF, London, November.

Rassenti, S.J., V.L. Smith and R. Bulfin (1982), "A combinatorial auction mechanism for airport time slot allocation", *Bell Journal of Economics,* 13, pp. 662-677.

SWEDEN

Jan Owen JANSSON
Department of Management and Economics
University of Linköping
Sweden

An analysis of the rail transport system

SUMMARY

Linköping, November 1997

1. BACKGROUND, PURPOSE AND APPROACH

As long as the national railways were fully vertically-integrated, "user charges for railway infrastructure" were not a big issue. Internal pricing of track services did not exist anywhere in the world.

Sweden was the first country to separate rail track planning and operation from the provision of rail transport services.

1.1. The Swedish experiment

What was the main reason for the creation of the Swedish rail track administration, "Banverket"? During the discussions that preceded the Swedish Parliament's decision to separate SJ and Banverket in 1988, the key sentence repeated over and over again was that "the road transport model should be adopted by the railways". This meant that rail transport infrastructure should be separately managed like the roads and be accessible to any train operators fulfilling strict safety requirements and paying rail track charges. The underlying reasons for this radical change were complex but it was clear at that time that SJ's chronic financial problems called for radical change; by dividing the organisation into two parts, it was thought that at least the new SJ, relieved of responsibility for the rail track, would sooner or later be able to pay its way, and also be stimulated by competition on the rail track to rationalise its operations. The entry of new train operators was envisaged at the time of the decision, although the legal prerequisites did not yet exist for it.

1.1.1 Has the "road transport model" worked for the railways?

Almost ten years have elapsed since the separation of infrastructure from operations and it can be observed that competition is still rather limited. SJ still has the sole right to run main-line passenger services. Regional branch line services are increasingly put out to tender by the regional (county) transport administrations, which keep services going by county council subsidies. SJ wins most of the contracts for regional and commuter train services.

Freight train operation is open to all, and a few small, independent train operators and one or two big industrial firms with their own rail transport systems are emerging. Since freight trains mainly run by night and passenger trains by day, so far there has been fairly little overlap of the two types of transport as far as the demand for rail track capacity is concerned.

The level of rail track charges has not been a barrier to new entrants because they have been rather low since the creation of Banverket; revenue from them has covered just a fraction of the total expenditure on rail track investment, repair and maintenance. It is now even being proposed by the committee of inquiry preparing the new Swedish transport policy that freight trains be completely exempted from rail track charges (Banverket, 1997a). Banverket is not obliged to cover costs and is financed by grants from the central government. In recent years, these grants have been relatively large because one of the priorities of current transport and environmental policy in Sweden is to bring about a renaissance of electrically-powered rail transport. Heavy investments in new rail track in order to make possible much higher train speeds are the main means to that end. Before the separation of SJ and Banverket, rail track investment had been lagging behind for decades because of SJ's chronic financial problems. The most important effect by far of the separation in 1988 has been a surge in rail track investment that no one could have dreamt of before. It should be emphasized, however, that this investment has not been financed by the rail transport sector but by the taxpayer. The Swedish Parliament has obviously been willing to subsidise rail transport even more than it was before the separation of SJ and Banverket. As mentioned, environmental concerns have been a main motive, but the good news some years ago that the restructured SJ would need less and less financial support and ought to be able to pay its way in the near future may also have had a positive psychological effect on the Swedish Parliament's attitude to the rail transport sector as a whole.

Some sceptics claim that the SJ's good financial performance in recent years is to some extent due to "creative accounting"; other people, including the present author, consider that SJ *should not* be obliged to break even. Part of the large subsidies to the rail transport sector as a whole should be directed to SJ in order to make possible a low-fare (peak-load pricing) policy that exploits the relatively low marginal cost of rail transport services.

1.2. Purpose of this paper

One thing that was never thoroughly discussed, neither in the preparatory stage before the separation nor during the formative years when Banverket was being built up and the "leaner" SJ was getting off the ground, was pricing policy for rail transport as a whole. It seemed too tricky an issue for the politicians to get to grips with. They seemed to rest content with the idea that rail and road infrastructure were now treated on an equal footing -- marginal cost pricing and cost-benefit analysis are applied to both -- and wishfully thought that SJ would soon break even and increase its market share at the expense of fossil fuel-burning road and air transport.

The purpose of this paper is to step in where the Swedish debate on rail transport pricing policy left off.

1.3. The pricing analogy between rail and road transport

The theory of rail user charges is still in its infancy. The situation is rather similar to that in the 1960s when the theory of road user charges was emerging. When the main pioneer in the field, Alan Walters, ventured to stick his neck out in US discussions of highway financing in the 1950s and 60s, his argument (see Walters, 1961), which today is generally accepted by both economists and highway engineers, turned the conventional wisdom of the time on its head. He argued that road user charges were far too low in urban areas and too high on many non-urban roads, whereas the highway cost allocation studies which dominated the field at that time concluded the opposite.

The longstanding clash of opinion between the proponents of average cost and marginal cost pricing seems to resurface in the new field of railway infrastructure user charges (Dodgson, 1995, Hylén 1995 and European Commission, 1996). This must be avoided. However, it is not enough just to embrace "the road transport model", as in the Swedish example.

The big problem, as the author sees it, is that the pricing policy of a rail transport system in which infrastructure and operations are separated, will still differ from that of the road transport system in crucial respects:

– Even if there are several competing train operators on one and the same line, they will never be as numerous as all the independent road users on the different links of a road network: monopoly, duopoly, or at best (or at worst, some would say) oligopoly, will still characterise rail transport systems in which infrastructure and operation are separated. This could result in serious problems of market failure if not handled with care. Needless to say, the problems of technical co-ordination are much bigger with rail than with the road.

– Related to the last-mentioned point is another crucial difference between rail and road transport services. The former are almost entirely provided in the form of scheduled public[1] transport (SPT), whereas private car travel and own-account road haulage and carriage for hire or reward dominate road transport.

The highly developed theory of optimal road user charges is quite involved because it is assumed that road user charges should (i) regulate road capacity utilisation in order to avoid costly congestion, (ii) contribute to improved road safety by covering the cost of accident externalities, (iii) implement the polluter-pays principle, (iv) internalise noise pollution costs, (v) cover the costs of the wear-and-tear caused by motor vehicles. In addition there may be a problem of capital costs: if optimal congestion tolls are set too low to recover the capital costs of roads, some other financial instruments have to be applied. It goes without saying that charges to internalise transport externalities like air pollution, noise and accident spillover cannot be used to finance road capacity expansion.

A proper user charge system should cover all these costs but in addition -- and this is the author’s main point -- if market failure is apparent at the rail service provision stage, rail user charges should take this into account too, given that the goal of net social benefit maximisation is the same as for road transport.

Perhaps these problems look more formidable than they actually are. After all, electrically-powered trains do not pollute and rail passenger transport is very safe. On the other hand, the noise problem, especially in built-up areas, is a serious one and the external accident costs generated by level crossings are considerable.

The main problem, however, is that rail transport is an activity with sharply decreasing costs at both the rail track stage and the transport service stage because of its SPT character. Optimal rail user charges would, in most places, make a very insufficient contribution to rail track capital, maintenance and repair costs.

1.4. Comprehensive cost analysis with a view to deriving optimal charges for rail transport infrastructure use

Let us not anticipate too much, but tackle the problems in a systematic way. This means that we should take an overall view of the problem of rail transport pricing. In the present state of theoretical knowledge, it is not advisable to consider the problems of railway infrastructure user charge services within a narrowly defined system.

In the following two chapters, the theoretical basis for a comprehensive rail transport cost analysis is laid down. In Chapter 2, the short-run cost structure is considered and, in Chapter 3, the long-run cost structure. This is followed in Chapter 4 by a repudiation of total cost allocation and half-baked long-run marginal cost approaches to costing and pricing. In Chapters 5 to 8, short-run price-relevant marginal costs are categorised, and short-run cost and output relationships are discussed. Special attention is paid to rail track wear-and-tear, accidents and the level and structure of the price-relevant costs of rail transport services.

Tentative conclusions are drawn in Chapter 9.

2. THEORETICAL BASIS

It goes without saying that optimal pricing cannot be based solely on cost information. Demand functions to match the price-relevant cost functions are required too. In fact, the strong emphasis on the supply, or cost-side, in this paper is more apparent than real, because in line with standard practice in transport economics, we widen the conventional cost concept to what is usually known as "the generalised cost".

2.1. Total rail transport costs

Rail transport users both demand services and supply their own time going to/from stations, waiting for the train and then travelling by it. In other services, this twofold role is normally dealt with by treating the quality of service as a demand factor. User inputs in the production of transport services, on the other hand, are assigned a monetary value and treated more or less in the same way as the various producer inputs for the purposes of appraisal of transport infrastructure investment. In transport demand analysis, the translation of quality of service into user time cost in order to obtain a "generalised cost" instead of just the monetary value has also been standard practice for several decades.

A similarly broad definition is appropriate for discussing optimal transport pricing. The natural production function is a transport system consisting of transport infrastructure carrying moving vehicles, terminals for change of mode and parking facilities for idle vehicles, transport vehicles with crew and passengers, and the price of system-external resources used by transport. The corresponding total transport system costs (TC) can be conveniently divided into four parts in the case of rail transport:

$$TC = TC_{inf} + TC_{train} + TC_{ext} + TC_{user} \quad (1)$$

TC_{inf} = total cost of rail transport infrastructure
TC_{train} = total cost of the train operator(s)
TC_{ext} = total transport system-external cost
TC_{user} = total time cost of rail users
Q = transport volume

It is clear that the ultimate determinant of each of these total costs is transport volume, Q.

By taking the derivative of TC with respect to Q, the social marginal cost (MC) is obtained, i.e. the sum of all marginal costs:

$$\frac{dTC}{dQ} = MC = MC_{inf} + MC_{train} + MC_{exr} + MC_{user} \quad (2)$$

2.2. The fundamental optimality condition for pricing policy

The total benefit (TB) of rail transport is obviously also a function of Q (together with a host of other variables which can be ignored for the moment). The goal is to maximise the net social benefit, i.e. the difference between the total benefit and the total cost. A first-order condition for a maximum is that the derivative of total benefit with respect to Q, i.e. the marginal benefit, is equal to the derivative of total cost with respect to Q, i.e. the marginal cost (MC).

$$\frac{dTB}{dQ} = \frac{dTC}{dQ} \tag{3}$$

Since the costs in this case also include passenger or freight time costs, the optimality condition (3) is not obtained by setting the price equal to the marginal cost, as is normally done. Instead we use the "generalised cost" (GC), i.e. the total perceived (by the users) sacrifice of another transport mode, including time costs as well as monetary outlays, that should be equal to MC.

$$GC = MC \tag{4}$$

Subtracting the real part of GC from both sides of (4), i.e. the monetary outlay, the price in the usual sense of the word remains on the left-hand side, and the right-hand side is written $MC - AC_{user}$.

$$\text{price} = MC - AC_{user} \tag{4a}$$

It is a well-established postulate of transport economics that, for a homogeneous collective of transport users, the average user cost (AC_{user}) is equal to the "private marginal cost", or real sacrifice of one transport mode for another by an individual user.

It is obvious from a comparison of (2) and (4a) that the right-hand side of (4a) is *not* identical with the social marginal cost, MC. There is a need for a separate designation: the cost-expression in question is called the "price-relevant marginal cost" or just the price-relevant cost (PC) for short. Unfortunately there is a problem of terminology: other authors on optimal transport pricing use -- either unthinkingly or just to comply with the time-honoured maxim that "price should equal the marginal cost" in first-best optimum -- the term "marginal cost" without a distinguishing prefix when referring to a cost expression of the kind represented by (4a) above. This can sometimes be confusing but provided that one is aware of the problem there should be no difficulty.

3. WHY IS OPTIMAL PRICING OF TRACK AND TRAIN SERVICES VERY SPECIAL?

While it is wise to bear in mind the simple, theoretical premise set out above, it is not much use for practical purposes. How do we go about estimating all the price-relevant cost components?

We would argue that sound economic cost theory, elementary railway engineering principles and common sense can take us some way towards the final goal of obtaining a complete picture of the price-relevant cost structure. Hopefully, this will be demonstrated in what follows. However, it must be remembered that the modern price theory of rail transport is still in its infancy. There are still a number of basic theoretical and methodological problems to solve, and we cannot solve all of them at one stroke. For the moment, it is primarily a matter of choosing the right theoretical direction.

We are at a theoretical cross-roads. One direction represents the adoption of cost accounting principles from industrial firms with a view to running the railways in as "business-like" a manner as possible. While the objective is a commendable one if it serves to enhance cost awareness, the means to the end would be counter-productive: like road transport systems, rail transport systems do not lend themselves to the cost accounting principles of manufacturing industry. The reasons why are developed below.

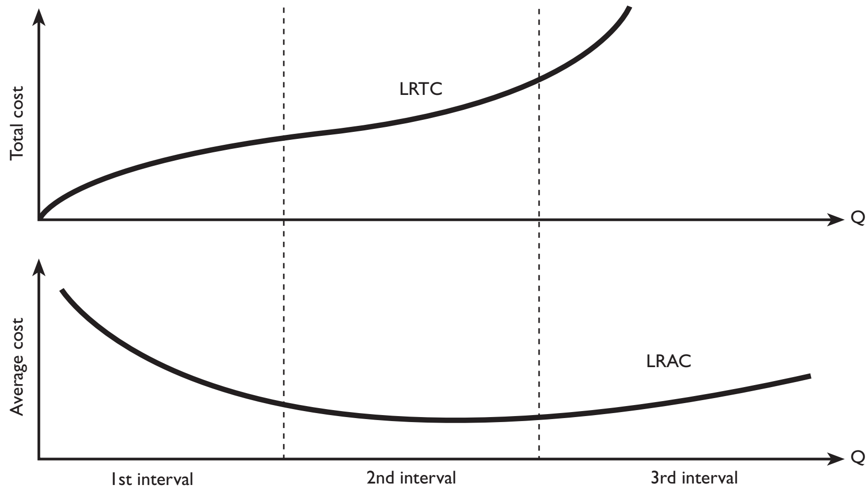
The other direction would take us along the line of development of the theory of road user charges, which started some forty years ago. We can learn a lot from this development, and if we can avoid repeating earlier mistakes, including the fruitless debate over short-run versus long-run marginal cost pricing, the theory of rail user charges will soon catch up with road pricing theory. These ideas are followed up in Chapter 4.

3.1. "The division of labour is limited by the size of the market"

The long-run cost structure is also relevant to transport pricing policy. It is not that long-run capacity costs should be added to short-run price-relevant costs -- that would be a sort of double-counting as will be explained later on -- but the long-run cost structure is crucial for first-best optimal pricing. It is the fundamental characteristics of the long-run cost structure of transport systems, coupled with the nature of the markets concerned that distinguish the price theory of transport infrastructure (as well as SPT) services from the price theory of mainstream microeconomics. The conventional cost diagrams in Figure 1 below provide a basis for a discussion of those characteristics.

The top diagram gives the *general* shape of the long-run total cost (LRTC) curve as a function of the output volume (Q) for *all* kinds of products, goods as well as services, in line with elementary production and cost theory (which is about as axiomatic as the notion that demand falls when the price rises). Initially, increasing returns-to-scale prevail, which means that the LRTC increases degressively with respect to Q. Thereafter, there is a relatively wide interval when basically constant returns-to-scale apply, but sooner or later decreasing returns-to-scale set in, and the LRTC increases progressively with respect to Q.

Figure 1. **The general shape of the curve of long-run total and average costs defines three cost-output intervals**



Needless to say, the character and extent of the three cost-output intervals can vary widely between different types of products. The general feature is that all three intervals exist for all products, which explains the (often rather drawn-out) U-shape of the long-run average cost (LRAC) curve in the bottom diagram of Figure 1. The flat part of the LRAC corresponds, of course, to the output interval in the left-hand diagram, where total costs increase in direct proportion to output.

Observations in the third output interval are difficult to make in reality, because a production plant with such a large capacity would be non-viable: the same output could be produced at a lower unit cost by two plants, each one with half the capacity. Actual observations could also be expected to be rare in the first output interval for a similar reason, at least as far as manufacturing industry is concerned. Small firms, or rather production at low-capacity plants of storable and transportable goods, could not survive if it is possible to produce such goods for substantially lower average costs at plants with higher capacity. Engineering cost studies of cost and output relationships, which examine technically feasible but not necessarily observable production scenarios, and consider plants of widely differing capacities with different technologies, regularly arrive at markedly L-shaped long-run average cost curves[2]. There is ample evidence in the engineering costing literature that small-scale production is often very uneconomic compared with large-scale production [see the classic, empirical studies by Haldi and Whitcomb (1967), and Pratten (1971) as well as a reference work like that by Sherer and Ross (1990)].

3.1.1. The nature of the product makes production under conditions of constant returns to scale atypical for transport infrastructure services

Mainstream price theory focuses on the second output interval, because it is assumed that most industrial production takes place in it. The price theory of transport infrastructure (TI) and scheduled public transport (SPT), on the other hand, should primarily focus on the first interval and, in the case of some TI services, on the third interval, because it is usually there that these services are produced. Why this is so can be explained as follows.

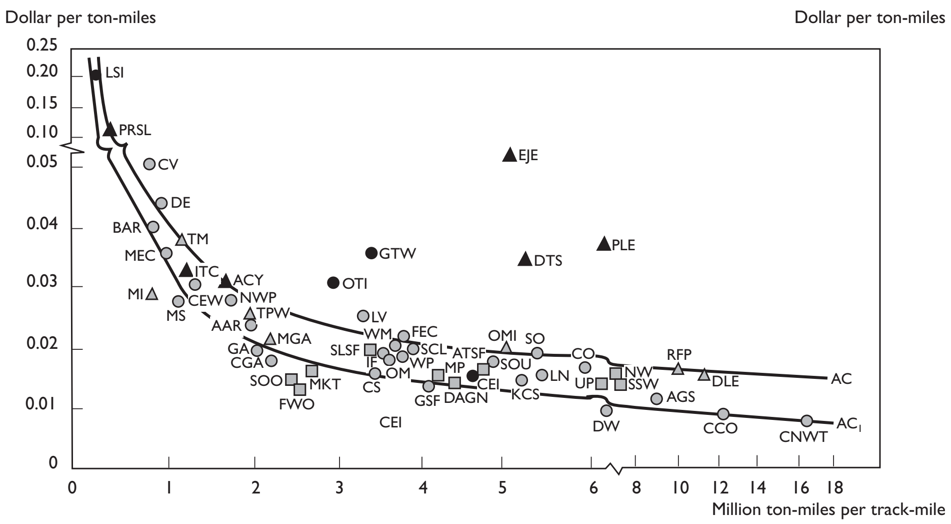
TI services and SPT services have the common characteristics of most services, i.e. they are non-storable and non-transportable and are "spatially unique" in the sense that transport between C and D is normally not a substitute for transport between A and B. This means that the number of separate, potential markets for TI and SPT services is very great indeed, which in turn means that the average market size is very small in terms of the demand base. The qualification "potential" is important. Bearing in mind the spatially unique character of transport demand, it can be argued that the most common market form as regards TI and SPT services is that there is no supply at all. The routes concerned are too thin and the average cost is too high to maintain any kind of TI, let alone direct SPT service.

This is amply confirmed by cross-section cost studies of transport infrastructure. Empirical observations are especially useful, because they make it possible to plot the entire long-run relationship between cost and output. The whole range of the L-shaped average cost curve will be represented among the observations. In manufacturing industries, the average costs of all plants and/or firms of a particular industry making easily transportable goods should be more or less on the same level, as a high-cost plant or firm could not survive where all plants in the industry serve the same national or even global market. Individual roads, on the other hand, are natural local monopolies. The fact that a particular motorway produces road services at a fifth of the cost per vehicle-kilometre of a small road between two villages in a different part of the country is of no consequence for the viability of the latter. The situation is different if the demand base is going to expand considerably, since it will sooner or later pay in terms of user cost savings to upgrade the existing road.

3.2. Marked small-scale diseconomies

There is little empirical data for rail track and train costs of the kind that exists for roads, airports and seaports. "Banverket" in Sweden has not existed long enough to produce reliable figures of the cost of rail track services with respect to traffic volume. The question of economies of scale in rail transport (rail track and train) services was once a hotly debated issue until it became clear that one has to distinguish firm-size economies and traffic volume economies on a particular line. Why were branch lines closed down when railway administrations were trying to improve their financial position? Obviously it was because costs per unit of traffic are typically several times higher than those on trunk lines, notwithstanding the high proportion of sunk costs in capacity costs. Harris (1977) pointed out that the question of firm size economies was not relevant to the issue of line closures, and showed that significant economies of traffic density exist in the US railway freight industry (Figure 2). The average costs comprise both track costs and freight train costs in this case. The article from which the figure is taken shows that more than half of the sharp decline in costs is due to marked small-scale diseconomies in the capital and maintenance costs of the rail track.

Figure 2. **Rail freight costs versus traffic volume in the United States**



Source: Harris (1977).

4. PRICE-RELEVANT COSTING *VS.* TOTAL COST ALLOCATION AND "LONG-RUN MARGINAL COST" APPROACHES

Modern price theory for TI services began with the theory of congestion tolls in the 1960s (Walters, 1961, 1968; Ministry of Transport, 1964), which applied the marginal cost pricing principle to urban traffic. Until now, the modern approach to pricing, consistent with the goal of social surplus maximisation, has coexisted with the traditional total cost allocation approach. The latter has its roots in American highway cost allocation studies from a long time ago; it boils down to allocating the "total cost responsibility" to each category of road traffic.

Total road vehicle-kilometres (**Q**) is composed of a number of traffic categories:

$$Q = \sum_{i=1}^{n} Q_i$$

Q_1 = total car-kilometres

Q_2 = total light truck-kilometres, etc.

The method of calculation in traditional total cost allocation is to first decide which parts of $TC = TC_{inf} + TC_{user} + TC_{ext}$ should be included in the "cost liability", which road users are required to pay in the form of motor vehicle and fuel taxation. In the second step the total cost liability is allocated between traffic categories 1 .... n in accordance with various criteria like axle load, etc. In the older versions of this approach, only Tc_{inf}, i.e. the total expenditure, or cost of the national road administration, was considered price-relevant. The main cost allocation problem was how to allocate road investment costs between cars, buses, light, and heavy trucks of different configurations. In later versions TC_{ext} is an important additional component which increases the total cost liability of the road traffic. The remaining problem with this approach is how to deal with the total costs of road users, TC_{user} This is not a road traffic-external cost to be sure, but should not the degree of congestion be reflected in the structure of road user charges?

The total cost allocation approach may at best result in a first approximation of the prices that are consistent with social surplus maximisation. To arrive at a more precise idea of those prices, we should focus on the *relationship* between the different cost components and **Q**. It should be stressed that the relationship between TC_{user} and **Q** is of vital importance for optimal pricing.

Marginal costing is not without problems, as attested by the long-standing debate on the relative virtues of short-run versus long-run marginal cost pricing.

4.1. Short-run *vs.* long-run marginal costing: why all the fuss?

When the marginal cost principle held sway in the 1960s, it was considered that road investment or capital costs were not price-relevant. Short-run variable costs had to be distinguished from the costs that were fixed in the short run, because it is the short-run marginal cost that was to be estimated and combined with the corresponding demand function to give the optimal price.

In retrospect, one can wonder why there was such a fuss about short-run versus long-run marginal costs? How could this issue dominate the debate, given that SRMC = LRMC along the growth path? Were there reasons to believe that there was systematic under- or overinvestment in the road network?

The reason why the debate could go on for so long is simply that the LRMC calculations which were presented as an alternative basis for road user charges were wrong, and therefore appeared to result in a much higher marginal cost.

4.1.1. The double-counting fallacy

One common mistake was to add infrastructure capital costs in the form of a “marginal capacity cost”, or something similar, to the short-run price-relevant marginal cost. Obviously, this meant that the short-run price-relevant cost was considerably increased. What was overlooked was the fact that, by increasing road capacity, more traffic can be accommodated without inflicting congestion costs on the original traffic. A trade-off is involved: when capacity-expanding investments are justified, an equivalent traffic (time, accident and vehicle operating) cost *is saved*, which would occur even if the investments were not made. Consequently, the additional traffic does not cause any congestion costs for the original traffic under these circumstances. A congestion cost component should not therefore be included in the long-run marginal cost calculation.

The mistake of counting both the additional capacity costs and the additional congestion costs that the capacity addition eliminates, was ascribable to the fact that the “square” total cost allocation procedure continued to be used.

4.1.2. Quality improvements ignored

Another even more common mistake was to forget that investments in the non-urban road network, where serious congestion is generally absent, are made primarily to *raise the quality* of road services (in other respects than eliminating traffic congestion). Let us consider this fact in the light of the empirical approach to long-run marginal cost estimation, known as “development cost” calculations. By means of time series analysis of the total costs of the whole interurban road network and the steadily growing total road traffic, it was thought that the indivisibility problem of long-run marginal cost calculations could be avoided: what appeared as a marked indivisibility on an individual road link was smoothed out when set in the context of the whole road network. The error of the "development cost" school of thought is, however, that one decisive term of the price-relevant long-run cost expression is overlooked. Only under the unrealistic assumption that road investments are purely capacity-expanding, without any accompanying quality-raising effect like shortening of distances or improved layout, would it be reasonable to take the ratio of the annual capital costs of new investments to the additional traffic on the roads as a proxy for LRMC. Since, on the contrary, non-urban road investments are almost wholly justified by cost savings for the existing traffic (including autonomous traffic growth), it is a big mistake to ignore the *negative* user cost term in the LRMC expression. The correct proxy for the price-relevant long-run marginal cost of the inter-urban road services has the following form:

$$LRMC_{proxy} = \frac{C_t - B_t}{\Delta Q_t} \quad (5)$$

C_t = Capital cost of new investments year **t**

B_t = Benefits for existing traffic year **t** in the form of user cost savings resulting from the new investments.

ΔQ_t = Addition to the traffic on the roads in year **t**.

It is well known that the traditional practice of national road administrations, when calculating the benefits of prospective investment projects, is to disregard newly generated traffic; only cost savings to existing traffic and projected autonomous traffic growth are included on the benefit side. Provided that the selected projects have a positive net benefit/cost-ratio, it follows that B_t is equal to or greater than C_t, which means that the price-relevant cost is zero or even negative.

There are, of course, in-between cases. In particular, in countries more densely populated than Sweden, it is presumably less rare that road investments (outside urban areas) are made, partly because there is a real shortage of capacity. In such countries, it is likely that the benefits to existing traffic, B_t -- given the rate of capacity utilisation -- will fall well short of the capital costs, C_t, and the value of the LRMC proxy will be positive, which is consistent with a positive congestion toll in this case.

The lessons that can be drawn from the long discussions of optimal charges for non-urban road services are generally relevant for all transport infrastructure, including railways. The boom in interurban railway investment in recent years was to a large extent justified by the resulting train user benefits in the form of straighter lines and, above all, higher speeds. Applying formula (5) for the $LRMC_{proxy}$ would most likely result in rather low marginal costs. There is much less urban (overground) rail investment in spite of the fact that there is strong demand pressure from commuters in many large cities. The big obstacles are land scarcity and encroachment costs. Once again, the problem for railways is similar to that for urban roads.

4.1.3. The importance of unaccounted-for encroachment costs for the relationship between short-run and long-run marginal cost

The equality of SRMC and LRMC presupposes that investments are optimal (defined by the goal of social surplus maximisation). The problem in practical cost-benefit analysis is that all the relevant costs are not readily translatable into monetary terms. The unaccounted-for costs may still have a strong influence on the relationship between SRMC and LRMC, in particular the encroachment costs of new infrastructure. It is very difficult to assign a monetary value to these costs and include them in cost-benefit analysis of transport infrastructure investments; they do, nevertheless, play an important role in the decisionmaking process. Many planned roads, railways and airports will never be built, or it may take a very long time to persuade politicians and local residents to accept a project in its original design. Financing difficulties are another frequent reason for infrastructure investments falling behind schedule, often compounded by the fact that the original projects are substantially modified in order to lessen the encroachment costs, which leads to additional building costs. This means that the SRMC and LRMC can be substantially different. As we should always use short-run costs for pricing purposes (disregarding sunk capital costs), we will focus on the SRMC in the following discussion. It is wise to bear in mind that the SRMC is *higher* than the LRMC under the aforementioned circumstances. The fact of discarding the long-run marginal cost will therefore not give a downwards bias to the result, as is often claimed by LRMC proponents.

5. THE SHAPE OF THE CURVES OF PRICE-RELEVANT COSTS FOR RAIL TRANSPORT

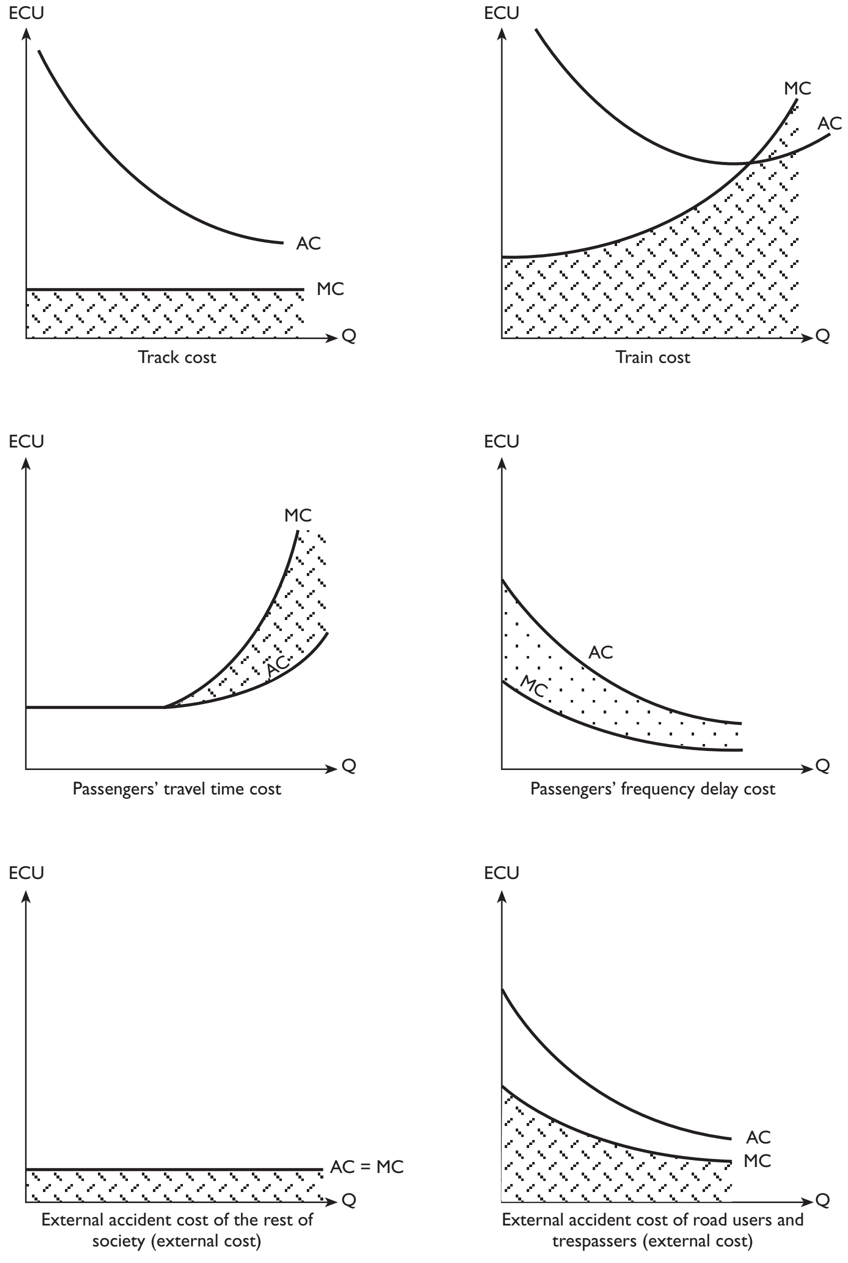
Having rejected total cost allocation and fallacious long-run marginal cost approaches as a basis for optimal pricing in rail transport systems, we shall now turn to the short-run cost structure. In the price-relevant short run, rail transport infrastructure is fixed, but like the labour input, the rolling stock on a particular line can be assumed to be variable, given the physical mobility of this type of capital input.

We shall consider track costs, train costs, user costs and external costs. For simplicity's sake, only passenger train services will be considered. The user costs then consist of passenger time costs, which can be divided into passenger time cost on the train, and so-called disguised waiting time costs, or the cost of "frequency delay" (a distinction first made by Panzar, 1979). We have included just one kind of system-external marginal cost -- the accident cost -- which is also divided into two: the external accident cost of road users and rail track trespassers, and the external accident cost to the rest of society. There are other costs of secondary importance which should, of course, be included in any calculation aiming at a complete picture of PC. The aim here, however, is to consider the main cost components and to examine some tricky aspects of rail transport pricing policy.

In Figure 3 below, the structure of the short-run cost and output relationships is given. The likely shape of the average (variable) costs and marginal costs (AC and MC) curves is depicted, and the price-relevant part is shaded (lined or dotted).

Track and train marginal costs and system-external marginal costs are fully price-relevant, whereas only the difference, MC--AC, is price-relevant to passenger costs. When AC is falling, which is true of the average cost of frequency delay, MC is below AC, and the price-relevant component is *negative*, i.e. a deduction in the total PC summation. This is the so-called Mohring effect ("discovered" by Mohring, 1972).

Figure 3. **The shape, in principle, of the short-run average and marginal cost curves for electrically-powered rail transport**



The cost curves drawn assume that rail transport capacity (engines and carriages) are optimally adjusted as passenger trips (Q) increase along a particular line. Disregarding fixed-formation trains, this will mean that trains get both longer and more frequent with increases in passenger demand. A simple optimisation model results in a "square root law": both the number of carriages per train and the number of trains (= departures) should increase in proportion to the square root of the transport volume along the expansion path for there to be an optimal trade-off between engine and engine driver costs and passengers frequency-delay costs. Interestingly, a recent comparison of the train input of Swedish Railways on different lines and line sections gave the following regression result (Molinder, 1997):

$$\text{Departures} = 8.51 \cdot Q^{0.48} \cdot e^{0.38D} \qquad (6)$$

where D is a dummy variable indicating whether or not a station at the end of a line section is a junction. The explanatory power of the simple equation is not greater than 61 per cent, but the volume-coefficient is highly significant (the t-value = 6.7) and, it can be seen, is very close to the 0.50 postulated by the "square root law".

This pattern of train capacity increasing along the expansion path also explains the assumption of a falling average accident cost (per rail passenger trip) inflicted on road users (at level crossings) and rail track trespassers. As a first approximation, the total accident cost can be assumed to be proportional to the number of trains. Since the number of passengers per train increases with the total transport volume -- roughly in proportion to $\sqrt{Q}$ -- the external accident cost per passenger falls as Q rises.

A fixed track capacity is the only reason why the train marginal cost, as well as the passenger travel time cost, will eventually rise, as depicted in the right-hand top chart and the left-hand chart in the middle row of Figure 3, respectively. This indicates that there is a profound, potential difference in the optimal price level of rail transport between some main lines with a generally high rate of track capacity utilisation, and branch and twig lines with a couple of short trains per day, as well as between peak and off-peak times on a particular line, capacity being strained at peak times.

While roads are common facilities, where a high rate of capacity utilisation leads to congestion and thus lower speeds, railways are better characterised as "departmentalised" facilities. When a train is using a certain section of the track, or "block", other trains cannot use the same block at the same time. If there is excess demand in the short run, queuing costs are incurred. In the long run, queuing costs should be provided for in the scheduling, as the planners of train timetables and itineraries will inevitably bear the consequences of scarce rail track capacity. "Scarcity costs" will translate into accelerating train and passenger time costs in one form or another as the rate of rail capacity utilisation rises.

An important point to bear in mind is that, if there is just one train operator in the system, the rail track scarcity costs are internalised in the timetable, which is made simultaneously for all train services in the network. There should be no need for rail track scarcity charges levied by the track owner on different trains. It is the task of the rail undertaking to design a peak-load pricing structure which optimises the rate of capacity utilisation over time, taking into consideration both rail track capacity and rolling stock capacity constraints. (This point is developed at more length in Chapter 8.) If this is not done satisfactorily, a complicated "second-best" situation occurs, where the track-owner may have to step in.

When a considerable number of independent rail transport companies use the same track, a system of rail track charges is advisable; this may be thought of as a set of "access charges" for blocks where excess demand would occur in the absence of such charges, so that the train operator with the highest willingness to pay gets first access.

5.1. Breakdown of the total optimal price under different organisational conditions

An algebraic formalisation of the preceding points is given below. This facilitates a more systematic treatment of how price-relevant cost should be divided between the rail track administration ("Railtrack" for short), the train service operator(s) and the Government, which levies externality charges or taxes on behalf of parties external to the rail transport system. The basic objective is that final consumers, i.e. train passengers, should pay the same total price-relevant marginal cost for rail transport, irrespective of how the rail transport system is organised.

First, the definitions of short-run average and marginal cost and output relationships are given in Table 1 below. It will be remembered that marginal cost is equal to average cost plus the product of output and the derivative of the average cost with respect to output.

Table 1. **Average cost and the difference between marginal cost and average cost**

Cost description	**AC**	**MC - AC**
Track wear and tear	f(Q)	$Q\frac{df}{dQ}$
Train operation	g(Q)	$Q\frac{dg}{dQ}$
Accidents	h(Q)	$Q\frac{dh}{dQ}$
Passenger time	t(Q)	$Q\frac{dt}{dQ}$

The total price-relevant marginal cost (PC), according to the definition in Chapter 2, is equal to the total marginal cost minus the user average cost -- passenger average cost in this case. The optimal fare should consequently be equal to PC.

$$\text{Optimal fare} = f(Q) + g(Q) + h(Q)$$

$$+ \underbrace{Q\frac{df}{dQ}}_{\text{negative}} + \underbrace{Q\frac{dg}{dQ}}_{\text{negative, positive}} + \underbrace{Q\frac{dh}{dQ}}_{\text{negative}} + \underbrace{Q\frac{dt}{dQ}}_{\text{negative, positive}} \qquad (7)$$

We shall look more closely at some components of PC in Chapters 6 to 8. The preceding reasoning, and the diagrams in Figure 3 indicate that MC_{track} and MC_{ext} are below AC_{track} and AC_{ext} , respectively, and MC_{train} and MC_{pass} can be both below and above AC_{train} and AC_{pass} respectively, depending on the rate of track capacity utilisation.

In a vertically-integrated rail transport system, there is only one price -- the train fare -- and an externality tax on trains, which will be included in the fare. In a separated system there are also rail user charges on trains, which, of course, will also be included in the fare. To illustrate our main point, let us disregard the process of converting rail user charges into train fares and instead look at the price structure as if Railtrack could charge passengers directly. Then we can regard all costs and prices as direct functions of passenger trip volume, Q. Table 2 below shows how the given total price of rail passenger transport should be divided between the track charge, the net train fare (excluding track and external costs) and the externality tax under the hypothetical conditions just stated.

It is assumed that Railtrack steps in and makes good the shortfall between the total price and the price-relevant cost. It could just as well be an agent of the Government -- a regulatory body or the Treasury, which in the present example is assumed to levy only an externality tax -- who does it. The point is that somebody with a good overview of the rail transport cost structure and the objective of net social benefit maximisation, has to step in when the market situation is such that deviations from the fundamental optimality condition would otherwise occur.

In Table 2, four different situations are illustrated. Further variants are, of course, conceivable -- for example, a variant between (3) and (4) where "competition among the few" prevails. Railtrack's compensating action would then naturally be somewhere in the middle of the bottom half of Table 2.

If total cost recovery is required in the rail transport system, it can already be concluded at this stage that this requirement would in most cases conflict sharply with the fundamental optimality condition. If the cost recovery requirement is imposed separately on Railtrack, a two-part structure of rail user charges would be advisable. If there is just one rail transport company, a fixed, lump-sum levy on the company that would allow Railtrack to break even would be preferable; "poll taxes" also seem the best solution when there are several train operators on the line.

Table 2. **Different distributions of the total price**

Rail transport system organisation	Total price		
	Track charge	**(Net) train fare**	**Externality tax**
(1) Vertically fully-integrated NSB-maximising concern	$f(Q) + Q\frac{df}{dQ}$	$g(Q) + Q\frac{dg}{dQ} + Q\frac{dt}{dQ}$	$h(Q) + Q\frac{dh}{dQ}$
(2) Separate Railtrack and single NSB-maximising train operator	ditto	ditto	ditto
(3) Separate Railtrack and single profit-maximising train operator	$f(Q) + Q\frac{df}{dQ} + \left[g(Q) + Q\frac{dg}{dQ} + Q\frac{dt}{dQ}\right]\frac{1}{1+e}$	$\left[g(Q) + Q\frac{dg}{dQ} + Q\frac{dt}{dQ}\right]\frac{e}{1+e}$	ditto
(4) Separate Railtrack and many profit-maximising train operators	$f(Q) + Q\frac{df}{dQ} + Q\frac{dg}{dQ} + Q\frac{dt}{dQ}$	$g(Q)$	ditto

6. PRICE-RELEVANT MARGINAL COST OF RAIL TRACK WEAR-AND-TEAR

The total expenditure of Railtrack (or "Banverket", when the Swedish administration is referred to) can be divided into four main categories: rail track *investment*, *reconstruction*, *maintenance* and unallocatable *overheads*.

The short-run marginal track cost, MC_{track} should be based on the wear-and-tear caused by rail traffic, which necessitates maintenance and reconstruction of track, bridges, catenary systems, etc. Major rail track investments are normally not justified when the existing infrastructure is worn out and/or is not worth repairing, because new track is shorter, has superior alignment, and allows higher train speeds -- in short, raises the quality of rail transport services. This will make track investments along existing routes profitable sooner or later when total traffic is on the increase (Banverket, 1997).

Table 3. **Breakdown of expenditure on rail track maintenance and reconstruction by Banverket in 1996 (%)**

Track-specific maintenance	**34**
remedying failures	3
damage	1
inspection	2
servicing of switches	2
replacement of rails and sleepers	4
adjustment of track position	2
periodic maintenance	1
winter maintenance	2
other maintenance	15
Non-track maintenance	**33**
Reconstruction	**33**
tracks	15
bridges	5
catenary systems	4
electrical equipment	4
signalling	3
marshalling yards	2
TOTAL	**100**

As Table 3 shows, Banverket's total expenditure on track maintenance and reconstruction which could be used as the basis for calculating the short-run

Mc_{track}, breaks down as follows: one-third on reconstruction, one-third on track-specific maintenance, and one-third on non-track maintenance.

The last third is thus excluded from the reckoning from the start.

6.1. Price-relevant maintenance costs

Of the other maintenance costs, some are fixed, or very nearly fixed, i.e. are independent of traffic volume (Hedström, 1996). Some activities fall into the category of "public goods", i.e. they benefit an unlimited number of trains; for example, cutting the grass and removing the weeds from railway embankments. A correlation analysis of expenditure on this kind of public goods-producing activity and traffic volume, would most likely show a significant relationship. However, this cannot be a basis for calculating Mc_{track}. As in expression (5), there is the possibility that the $LRMC_{proxy}$ could take a value of zero: while total expenditure or cost (C) is found to rise in proportion to traffic volume (Q), it would be hasty to conclude that price-relevant marginal cost is equal to average cost. It may be that some offsetting benefit (B) is also increasing with C and Q. In the investment case discussed in Chapter 4, it was pointed out that the cost savings for existing traffic typically remain in line with C over time, making the long-run marginal cost of non-urban transport systems close to zero. In the present case, it can be established that a number of qualities of rail track services are enhanced with increases in traffic volume. In order to calculate the relevant MC_{track} component, the correct procedure in this case is also to deduct the user cost savings from the producer cost, which would typically nullify the price-relevant cost (see further, Jansson and Lindberg, 1997).

6.2. Price-relevant reconstruction costs

Replacement of components of existing rail infrastructure constitutes another third of Banverket's total expenditure, excluding investment expenditure proper. As Table 3 shows, track (rails and sleepers) is by far the single biggest item. The method of calculating the price-relevant marginal cost for this is first to estimate how the interval between replacements is affected by traffic volume, i.e. to estimate the useful life of different components with respect to traffic volume. Then the present value of total reconstruction costs is calculated as a function of useful life, and the price-relevant reconstruction marginal cost is obtained by derivation with respect to traffic volume. For example, if the useful life is inversely proportional to traffic volume, the price-relevant marginal cost is equal to the average reconstruction cost. On the basis of engineering rules of thumb, it would seem that the useful life of most components (rails, sleepers, electrical equipment, catenary systems, etc.) is not that closely dependent on traffic volume.

A safe guess would be that the price-relevant short-run MC_{track} is less than half the average cost of railway infrastructure maintenance and reconstruction.

The structure of MC_{track} with respect to axle load is a different matter. First, it should be pointed out that track quality strongly influences the cost of wear-and-tear. In the same way as for roads, there is a trade-off between investment in quality and the necessary maintenance cost, with the traffic volume as the balancing factor. The least well-constructed track will thus have the highest marginal wear-and-tear costs -- some evidence suggests about five times higher (on Swedish type II track) than on high quality track. Wear-and-tear is also dependent on axle load for a given traffic volume. Those studies which have tried to establish a direct link between maintenance cost and axle load have arrived at an almost linear, or slightly progressive, relationship. Although it is not possible to draw any firm conclusions about how reconstruction and maintenance costs increase with axle load, it is clear that the relationship is far less progressive than the so-called “fourth power law” in the road sector.

7. PRICE-RELEVANT MARGINAL COSTS OF ACCIDENTS

In analysing accident costs, a basic distinction needs to be made between accidents that involve only rail users, and accidents that occur at interfaces with other transport systems. Rail-road crossings are notorious accident black spots (see Banverket, 1997, Evans and Morrison, 1997 and Johansson, 1997). It may also be noted that a large proportion of serious road accidents occur where pedestrians and cyclists have to cross roads used by motor vehicles. The total number of accidents, N can be written:

$$N = A + X, \qquad (8)$$

where A stands for the number of intra-system accidents, and X for inter-system accidents. As may be suspected from the data in Table 4 below, it is the X-accidents that give rise to the really significant component in optimal accident externality charges in rail transport systems.

There are many theoretical and empirical problems involved in the valuation of the more serious types of accidents. A brief summary of some relevant points is given below:

7.1. *Ex ante* and *ex post* accident costs

Total *ex post* accident costs are literally "inestimable". *Ex ante* costs, on the other hand, can be defined as the total willingness to pay (WTP) for complete safety, on the part of different individuals directly or indirectly exposed to accident risks. We can distinguish the WTP of three groups:

$f(r)$ = willingness to pay for complete safety, of the household to which a certain person exposed to the risk r belongs; $f(r) \approx ar$ for low values of r ;

$g(r)$ = ditto, of relatives and friends of the above person; $g(r) \approx br$ for low values of r ;

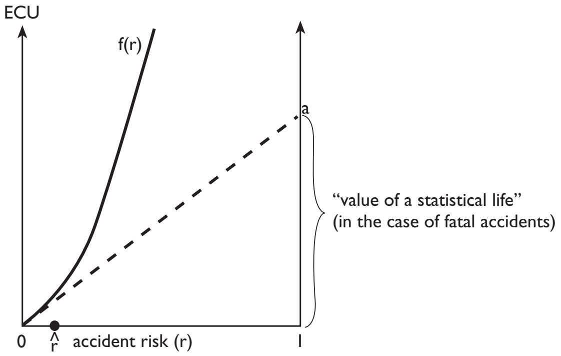
$h(r)$ = ditto, of the rest of society; $h(r) \approx cr$.

Table 4. **Rail accidents in Sweden 1989-96; number of persons killed or seriously injured (including suspected suicides)**

Accident category	**Killed**	**Seriously injured**
Railway staff	***7***	***44***
hit when working on the track	5	12
(ranging work)?	2	10
level crossings	0	0
other	0	22
Passengers	***10***	***68***
level crossings	2	6
other	8	62
Other persons	***166***	***100***
road users at level crossings	127	88
trespassers	39	12
TOTAL	***183***	***212***

Source: SJ (1997).

Figure 4. **Households' willingness to pay, at different risk levels, for complete safety for a particular household member**



The assumed general shape of f(r) is depicted in Figure 4 above. It can be assumed that this function is approximately linear for very small risks (below $\underline{r}$) coinciding with a line from the origin, **ar**, which obviously takes a value = **a** at a risk level of unity. At high risk levels, however, the tip of the shape of the function f(r) bends upwards at an accelerating rate, and will never reach the range close to **r**=1.

The function g(r) presumably (but not necessarily) rises less steeply than f(r), and h(r) could be approximated to **cr** in the whole range; it represents the "cold-blooded" part of the total accident costs, i.e. the assumption that the rest of society regards victims of traffic accidents in the same way as hard-hearted slave-owners regarded slaves.

The main difference between *ex ante* and *ex post* valuations of serious casualties should be noted: let us assume that **r** stands for fatality risk, and that a representative household is prepared to pay f(r) (≈ **ar** for low risks) for complete safety for a particular household member. If 1/r households have a similar willingness to pay for safety, it can be concluded that one life will be saved for a total payment of f(r)/r ≈ **a**. However, no normal household would be willing to sacrifice a member for a sum of money = **a**; life is invaluable. Compensation to families which have lost a dearly-loved member can help to mitigate their grief and suffering; the same goes, of course, for disabled victims of traffic accidents. It should be clearly stated that there is no economic way of determining the right compensation. The "value of a statistical life", **a**, for example, is an arbitrary entity; a sum of money to compensate bereaved families is equally arbitrary.

7.2. From *ex post* cost responsibility to *ex ante* externality charges

In a vertically-disintegrated rail transport system, the responsibility for accident costs can be shouldered by the track owner or the train operator, or both. The governing principle should be that the organisation(s) which oversee the safety of rail transport should have an effective incentive (in the form of *ex post* responsibility for accident costs) to improve safety. In any case, these costs will be passed on to rail transport users, paying the corresponding *ex ante* charges as part of train fares or freight rates. The question is: how is the price-relevant marginal cost to be calculated?

Let us disregard the tricky matter of the division of the *ex post* cost responsibility between the track owner and train operator(s), and assume the point of view of a vertically-integrated railway undertaking which seeks to maximise the social surplus: how should the fact that both intra-system and inter-system accidents are by-products of the transport services provided be taken into account in the pricing of train services? The following principles seem consistent with the goal of social surplus maximisation.

The railway undertaking pays a legally stipulated sum **a + b + c** in compensation for every intra-system casualty, of which **a** goes to the households of the victims, and **b + c** to the public purse. (A small company with modest turnover would have to take out insurance against serious accidents.) To get from these *ex post* expenditures to a price-relevant accident cost which can be included in the optimal rail fare, the relationship between the number of accidents (A) and transport volume (Q) first has to be estimated.

$$A = f(Q) \tag{9a}$$

The accident risk, **r** is defined:

$$r = \frac{A}{Q} = \frac{f(Q)}{Q} \tag{9b}$$

The fact that the railway undertaking bears full *ex post* responsibility for accident costs is a strong incentive to improve safety. Transport users can do little about this apart from abstaining from less essential trips. For the trade-off to be consistent with the goal of social surplus maximisation, the marginal expected accident cost has to be included in the train fare. Transport users become aware of risks, or rather the cost of accident risks, through the optimal fare. The intra-system accident cost component in the price-relevant cost is derived as follows:

$$PC_A = \frac{dA}{dQ}(a+b+c) \tag{9c}$$

$$= (1 + E_{rQ})\frac{TC_A}{Q} \tag{9d}$$

If the number of accidents, A, is proportional to the traffic volume, Q, the risk-elasticity, E_{rQ} is zero, and the accident charge equals the average intrasystem accident cost. It should be remembered, however, that a good part of the revenue from this charge goes back to transport users/payers and their families in the form of reimbursements of the costs of medical care and sickness benefits and compensation for grief and suffering, assuming that the general health insurance scheme does not cover these costs. However, in rail passenger transport, average intra-system accident costs are rather small. Inter-system accidents are a somewhat different matter.

Unfortunately, trains occasionally hit cars, cyclists, or pedestrians. The outcome of such collisions is nearly always that only the lighter vehicles are seriously damaged.

The number of inter-system accidents is X, which is a function of both train transport volume Q, and the traffic volume of the other transport system, denoted by M.

$$X = g(Q, M) \tag{10a}$$

The inter-system accident cost component in the price-relevant cost is obtained as follows:

$$PC_x = \frac{dX}{dQ}(a + b + c) \tag{10b}$$

The risk for the other party of such accidents is denoted by R. Since R = X/M, we can alternatively write PC_X like this:

$$PC_x = \frac{dR}{dQ}(a + b + c)M \tag{10c}$$

$$= E_{RQ}\frac{TC_x}{Q} \tag{10d}$$

To sum up, the total price-relevant accident cost comes to:

$$PC_{A\&X} = (1 + E_{rQ})\frac{TC_A}{Q} + E_{RQ}\frac{TC_X}{Q} \tag{11}$$

On the one hand, intra-system and inter-system average accident costs are factors of potential importance and, on the other, the risk-elasticities with respect to passenger travel volume, E_{rQ} and E_{RQ}, are important for the price-relevant accident cost. What is the likely order of magnitude of $PC_{A\&X}$?

7.2.1. The order of magnitude of price-relevant accident costs: a numerical example based on Swedish data

In Sweden, the position is as follows: the intra-system average fatality and serious injury accident cost, TC_A/Q, counting both passengers and railway staff, is no more than 1.5. It was 10^{-3} ECU per passenger-km, or .3 ECU per average train trip of 180 km during the period 1989-96. The value of the risk-elasticity, E_{rQ} is not known but is unlikely to differ much from zero. Passenger casualties could be expected to be proportional to transport volume, and staff casualties proportional to the number of rolling stock units in operation, which in turn is roughly proportional to transport volume. The proportionality between A and Q means, of course, that the risk (r) is constant with respect to the volume of travel, and that the risk-elasticity is zero.

The corresponding inter-system accident cost, TC_x/Q is many times higher. In Sweden it was $7.5 \cdot 10^{-3}$ ECU per passenger-km during the period 1989-1996 if trespassers are included and $6 \cdot 10^{-3}$ ECU per passenger-km if they are excluded. The risk-elasticity E_{RQ}, i.e. the relative change in the risk for external parties as train passenger volume changes is, like E_{rQ}, strictly unknown, but common sense suggests a value between 0 and 1. Given the number of road vehicles crossing railway lines, the more trains there are, the greater the risk to road users of a collision. It is unlikely that the number of collisions will increase proportionally to train passenger volume Q. As was pointed out in connection with the cost-output diagrams in Figure 2, train lengths will increase as Q increases, which means that the number of trains will not increase in the same proportions as Q.

It is worth noting that, for traffic on urban roads, a principal component of the price-relevant accident cost is the cost inflicted on unprotected (vulnerable) road users, mainly at junctions and at pedestrian crossings. However, this price-relevant cost is only about half the average cost of accidents involving protected and unprotected road users, because the corresponding risk-elasticity seems to be around one half (Brüde and Larsson, 1993), i.e. given the volume of travel on foot and by bicycle, the number of pedestrians and cyclists hit by vehicles increases in proportion to the square root of the motor vehicle traffic volume.

It is quite possible that E_{RQ} is also about one half in the present case of collisions at level crossings. The total price-relevant accident cost for rail passenger transport in Sweden could consequently be around $6 \cdot 10^{-3}$ ECU per passenger-km, or close to 1 ECU per passenger trip of average length.

8. PRICE-RELEVANT MARGINAL COST OF RAIL TRANSPORT

Given the externality charges and rail track user charges levied by Railtrack and/or the Treasury, what is the optimal passenger fare? This question forms a large part of transport economics. Here we shall only make a few simple, but nonetheless fundamental, points. The basis for the discussion is that fares should be equal to the marginal train cost plus the difference between marginal and average train user costs.

$$PC = MC_{train} + Q\frac{dAC_{pass}}{dQ} \qquad (12)$$

The number of trips by the rail transport system concerned is Q. The external marginal cost appears as the charges payable by the rail undertaking. These charges are consequently part of MC_{train}, and will be passed onto rail passengers in the fares they are charged.

It should be pointed out that each of the two price-relevant cost components of (12) will take quite different values depending on the way in which additional passengers (or freight) are taken on. The sum of the two components is the same in an optimum situation, irrespective of how capacity is augmented. Three examples illustrate this point:

– Additional passengers can normally be accommodated (or almost) without any additional producer inputs, simply by increasing the occupancy rate. During peak periods in particular, this cannot be done without having an effect on passenger costs. PC will in this case consist solely of an occasionally high passenger cost component, representing the queuing and/or crowding costs of passengers or freight.

- Another, more frequent, way of accommodating additional customers is to increase the number of trains. In this case MC_{train} will be fairly substantial, while the user cost component becomes negative due to the Mohring effect.
- A third possibility is to increase the size of vehicles -- train length in this particular case. This would leave the passenger cost component in the price-relevant cost more or less unchanged, and only MC_{train} would contribute to PC.

8.1. The average cost of a marginal carriage

The theory lends itself particularly well to practical calculations in rail transport, where train size (length) is adjustable; carriages can be added to or uncoupled from the train in a marshalling yard during the night. Boarding/alighting charges are, moreover, a very minor complication because, unlike with bus transport for example, the number of inlets and outlets increases proportionally to vehicle size and tickets are bought in advance, which means that the ticket transaction time is not part of the transport vehicle time.

It is more difficult to estimate the "Mohring" effect in long-distance public transport, for which prospective travellers use timetables, than in urban transport, including commuter rail services. The right approach for long-distance rail transport, therefore, is to calculate PC by assuming that additional passenger demand is met by increasing vehicle size. This approach avoids most of the difficulties of passenger cost estimation, but not entirely. If train length increases proportionally to the number of passengers, it should in theory be possible to keep total time at stations constant, irrespective of the number of carriages. In practice, this time increases slightly with the number of passengers. However, the boarding/alighting charges should play a relatively minor role.

In calculating the incremental cost of adding another carriage to a train, it is also necessary to find out how energy costs change as a train is lengthened. In a joint study with SJ, a Linköping University research team found that, for a given train speed, energy consumption will increase linearly with train length in the whole range of observations (Jansson *et al*., 1992). The price-relevant cost can then be formulated in a very simple way. In the normal scenario, the least unit of supply is another carriage carried from the point of departure, say, the central station of Stockholm, to the final destination, for example, Malmö, and back again. The incremental cost of producing this additional capacity constitutes the numerator of the pricing-relevant cost, and the number of additional passengers thus accommodated constitutes the denominator:

$$PC_{ti} = \frac{\mu_{ti} + \alpha D}{\beta} \qquad (13)$$

PC_{ti} = pricing-relevant cost per occupied seat day t train departure i (t = 1..... 365, and i = 1 ......m)

μ_{ti} = opportunity cost day t train departure i of the marginal carriage

α = additional running cost of a train per kilometre caused by coupling up another carriage

D = round trip distance

β = target number of occupied seats per carriage.

This formulation of the price-relevant cost presupposes that the train on the route concerned only makes one round trip per day. On a shorter route, it may be possible to make one and a half or two round trips, which would mean that the denominator has to be increased by a factor of 1.5 or 2.

Note that the price-relevant cost is given per occupied seat on a round trip. This cost should be shared out among all the passengers successively occupying a particular seat during a round trip. The number of passengers per seat and round trip could be two, one in each direction, or more than two, since many passengers do not travel the whole distance.

An efficiency condition is that, summed over all the departures every day of the year, the opportunity cost of a carriage should equal the annual capital cost.

The financial result of optimal pricing of passenger train services is easily imagined. The revenue will cover the capital and operating costs of carriages, including guards' wage costs, but will make no contribution to the costs of engines, including engine-drivers' wage costs, nor to the major part of overhead costs, which are independent of train length. Only about half the total costs of passenger services will be covered by optimal train fares.

8.2. Peak-load pricing of interurban train services

With reference to expression (13) for the level of the price-relevant cost of rail passenger transport, an additional efficiency condition is useful for the derivation of the peak-load pricing structure. It can be written as follows:

$$\mu_{t1} = \mu_{t2} = ... \mu_{ti} ..... = \mu_{tm} = \mu_t \qquad (14)$$

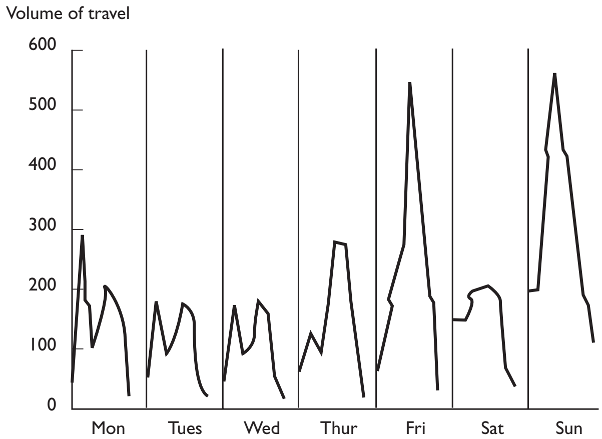
The rolling stock on a particular line can be assumed as given on one particular day. The number of engines and carriages can only be changed from one day to another. An efficiency condition is then that each day the given number of carriages should be distributed between the *m* trains such that capacity utilisation is nearly constant. This means in turn that the opportunity cost of a carriage is the same for every departure on a particular day, as shown in (14) above.

The stochastic element in rail travel demand is substantial, so a very high occupancy rate should not be aimed at. The mean occupancy rate of SJ's trains is currently about 1/3 but it varies systematically in different sub-markets. By eliminating the systematic differences by means of peak-load pricing, aiming basically at equalisation of the train occupancy rate in time and space, it should be possible to raise the mean occupancy rate to at least 1/2, which would be a very considerable improvement.

The first demand equalisation to aim at should be to make the Monday-Thursday and Saturday (off-peak) level of demand nearly equal to the Friday and Sunday (peak) level. A representative example of the time profile of rail travel demand by day of the week in Sweden is given in Figure 5.

It was found in Jansson *et al.* (1992) that $\mu i = 0$ at off-peak times, i.e. if fares at off-peak times were based on just the running cost component α D of the price-relevant cost [see equation (4) above], the level of demand would just fall short of the peak level as it would do when peak traffic alone pays the carriage's capital costs.

Figure 5. **Rail travel between Stockholm and Gävle on different days of the week**



The second demand equalisation to aim at should be to level out spatial peaks and troughs. We have not gone into this matter very deeply and a lot still remains to be done. We shall take just a rather typical example -- the line between Stockholm and Sundsvall in the north of Sweden. Dividing it into three sections, the daily passenger flow at peak and off-peak times on the three sections was as follows:

Table 5. **Passenger flow per day on three sections of the Stockholm-Sundsvall line in 1984**

Line segment	**Fri, Sun**	**Mon-Thur, Sat**
Stockholm - Gävle	3 307	1 751
Gävle - Söderhamn	2 490	1 240
Söderhamn - Sundsvall	1 494	734

The peak/off-peak price differentiation advocated above would equalise the flow figures in each row of Table 5. To match spatial supply and demand better, the first step is to make some trains from Stockholm turn around in Gävle, and some in Söderhamn. Spatial demand equalisation would then imply a price differentiation with the aim of making the occupancy rate equal on each section of the line.

8.2.1. Comparison with SJ's fares

The following comparison of SJ’s fares and optimal fares according to the principles laid down in this report focuses on fare differentiation by day of the week (see also Jansson *et al.,* 1992). As can be seen in Table 6, in both the cases investigated (with and without a budget constraint), off-peak fares should be only about one-third of the peak fares. Examples of fares are given for a rather short, a medium-distance and a fairly long-distance route. The absolute values of the figures -- in 1990 Swedish kronor -- can be disregarded. It is the fare structure that is interesting. Since the elasticity of the demand for rail travel differs somewhat for routes with and without airline competition, both cases are considered in the illustration.

It can be seen that the distance makes little difference as far as the peak/off-peak differentiation is concerned. It is worth noting that where air transport is not an alternative, the distance-dependency is about the same as in SJ’s tariff, and that where air transport is an alternative, fares taper off more markedly in both cases of optimal fares.

The comparison with SJ’s fares is complicated somewhat by the fact that SJ has abandoned the differentiation of fares by day of the week, which was the basic feature of its 1979 "low-fares policy". In 1989, it introduced a new fare differentiation, with reduced fares for departures in the same day. Some 20 per cent of departures every day are "red", which means that the fares are only half the fares that apply to the remaining 80 per cent of "black" departures.

Table 6. **Comparison of optimal rail fares for different days of the week and SJ's fare structure in 1990, SK per second-class single trip from start to end**

Distance	Day of the week	Optimal fares				SJ fares	
		without budget constraint		with budget constraint			
		Air competition	No air competition	Air competition	No air competition	Black	Red
170 km	Fri, Sun	113	113	154	154	128	64
	Mon-Thu, Sat	30	30	56	57		
335 km	Fri, Sun	187	202	240	303	242	121
	Mon-Thu, Sat	50	58	87	123		
550 km	Fri, Sun	254	294	292	462	342	171
	Mon-Thu, Sat	72	96	104	204		

In autumn 1997, SJ reverted to a peak-load pricing structure by day of the week. It has partly copied the fare structures of airlines. It now combines proper peak-load pricing and price discrimination in a way that is not quite clear to the general public (which is also the idea). Peak-load pricing does not apply to first-class passengers, who are mainly business travellers who do not pay for their tickets out of their own pockets. Second-class seats are divided into two categories -- full-price and half-price. There is no visible difference between the two categories but their relevant proportions vary widely on different days. On very busy days (Fridays and the eve of departures, etc.) there are hardly any half-price seats but on days when demand is slack there are plenty of them. To prevent cost-conscious business travellers from taking advantage of the low fares, two conditions have to be met: (a) an annual pass at the modest price of 18 Ecus has to be purchased and (b) tickets must be booked one week in advance.

If social surplus maximisation were the goal of SJ, i.e. maximisation of the sum of the (positive or negative) profit of SJ and the rail consumers' surplus, this kind of peak-load pricing would need to be open and implemented in full. The low off-peak fares in the optimal tariff structure would apply to 70 per cent of total travel, which means that the weighted average fare level would be substantially lower than the level of SJ's fares in the Table, even when the financial result is constrained to the same extent as that of SJ in the comparison.

We made a rough calculation of the likely increase in transport volume as a result of adopting the optimal fare structure in full. In the unconstrained case, transport volume would double. Most of the increase would, of course, occur in the off-peak period. In the constrained case, the price level has to be substantially higher; as a consequence, the increase in the total volume of travel falls to 40 per cent. It is worth noting that the net welfare gain in the latter case is as high as 75 per cent of the net welfare gain from peak-load pricing in the case where no budget constraint is assumed.

9. TENTATIVE CONCLUSIONS

We are in unexplored territory, so the following summing-up is necessarily "tentative". The salient economic feature of rail transport is the pronounced economies-of-demand density. Only on the very busiest inter-city lines and, in particular, urban commuter services, will physical barriers to capacity expansion be sufficiently significant to turn rail transport into a constant- or increasing-cost activity. Pricing and investment policies are mutually dependent on one another. If rail track investment falls behind for a long time, the number of bottlenecks on the railway network will increase, taking the form of increasingly frequent delays and making it increasingly difficult to schedule services, with forced detours and waiting times. In other words, at a suboptimally low level of investment, it is theoretically possible to evoke the increasing-cost characteristics of rail transport; among other things, this would imply that optimal pricing policy is consistent with total system cost recovery -- at least for a certain period of time before demand starts to fall off, due to the decreasing quality of service.

Nobody should desire such a scenario. Instead, optimal railway investments should be continuously undertaken to eliminate bottlenecks, shorten rail connections and improve the track to make comfortable high-speed train services possible. Then, one can be certain that the total price-relevant cost of rail transport will fall short of the total average cost. Optimal pricing would not cover the total costs of infrastructure, traffic operation (including rolling stock capital costs) and externalities.

This may be considered a big problem in some countries. In other countries it seems acceptable to subsidise the railways, provided that the travelling public/taxpayers get good value for their money: whatever organisational form is chosen, the railways must acquire the image of efficient undertakings, responsive to the needs of their customers and charging reasonably low fares so that ordinary people can afford to travel by rail on a non-exceptional basis. Otherwise, subsidisation will be increasingly unpopular.

9.1. What really matters

The preceding analysis of the structure of rail transport costs in general, and track costs in particular, leads to the conclusion that there are no more dramatic changes that have to be made to improve allocative efficiency, except one: somehow or other, train operators need to be induced to adopt proper *peak-load pricing* instead of the traditional, distance-based fare structures, or so called "market pricing", which in fact boils down to skewed price discrimination. Making business travellers pay much more for the same service, except for relatively small comforts like meals served at your seat, is probably a good compromise between equity, efficiency and financial requirements. The important thing, though, is to make rail travel once again the popular mode of transport it used to be.

This is both possible and consistent with economic efficiency conditions. In off-peak periods -- roughly Monday to Thursday, and Saturday -- as well as on the return-haul on temporarily unbalanced routes and on line sections with a substantial amount of slack in demand, the price-relevant costs of rail transport are at a level which would make rail transport very competitive if the fare structure reflected those costs. More than half the population does not work regularly. Pensioners, the unemployed and students represent a large potential market for off-peak travel if only fares were set at the level of the price-relevant costs.

9.1.1. The price structure is more important than the price level

It is important to note that, even if a strict budget constraint is imposed on traffic operations, a peak-load fare structure will increase the social surplus substantially more than time- and space-constant fares.

The best alternative is probably social surplus maximisation, taking the so-called "shadow price" of public money into account. In Sweden, the costs of tax-financed public investment in the transport sector are inflated by a factor of 1.6 to take account of the fact that value-added tax is not levied on the services in question and, secondly, of the distortions caused by raising additional income tax revenue. The theoretical and empirical foundations of the current shadow price of public money are not very robust. The basic idea is that progressive income taxation is a wasteful, but for distributional reasons, necessary form of taxation. (To reduce the need for income taxation yields 1:30 for every nominal krona saved?) Apart from poll taxes and other lump-sum taxes, which are very unpopular, well-designed commodity taxes seem to be the least distorting form of taxation, and should be used as far as is politically possible, at least up to a VAT rate of 30 per cent, so as to reduce the need for income tax.

This means that a value-added-tax of this order of magnitude should be imposed on all public services sold in the market, irrespective of the total financial result. It is to the margin that the shadow price of public money should be applied.

9.2. The role of Railtrack in rail transport pricing policy

The problems and possibilities identified in this paper have now been summarised. The final, important question remains: what can Railtrack do about them?

In a vertically-disintegrated organisation, there should still be somebody who has an overall responsibility for the rail transport sector. A precedent is the Swedish Road Administration, which was recently given responsibility -- within its obviously limited capacity -- for ensuring that the development of the whole road transport sector is in line with the public interest. A similar role may be envisaged for Railtrack. Given the specific characteristics of supply (cost) and demand in the rail transport sector, Railtrack user charges should be used to compensate for deviations from the fundamental optimality condition stated at the outset.

In a country like Sweden, the potential for this is rather limited because the present level of track user charges is very low. While this is right from the point of view of allocative efficiency, it leaves little scope to press for peak-load pricing of train services.

In countries where cost recovery is a requirement imposed on the Railtrack authority, the potential to influence the final price structure of rail transport is, of course, greater.

It is difficult to generalise about how the rail track administration should perform this balancing act. If there is just one train operator on each particular line, by far the best approach is to try to make the goals of the administration and the train operator the same, i.e. social surplus maximisation. Then there should be no need to compensate for deviations from first-best optimal pricing of Railtrack services.

If there are several train operators on the same railway line, maximising individual profits, a number of deviations from the efficiency conditions could easily occur. It is difficult to conceive that a Railtrack pricing policy would be sufficient to compensate for all deviations.

NOTES

1. The term “public transport” is not used here to mean that transport services are publicly produced and/or that the service producer is a public enterprise, but that the transport vehicle in question is used by a number of independent transport consumers in common, i.e. at the same time on the same route.

2. The decreasing returns-to-scale in the third output interval, necessary for the commonly assumed U-shape of LRAC, which are caused by organisational "diseconomies of size" rather than technical limitations, are more difficult to verify by theoretical engineering calculations.

BIBLIOGRAPHY

Brüde, U. and J. Larsson (1993), *”Trafiksäkerhet i tätortskorsningar”,* Linköping, VTI-meddelande 685.

Dodgson, J. (1995), Railway privatisation and infrastructure charging, *7th World Conference on Transportation Research*, Sydney, July.

European Commission (1996), *A strategy for revitalising the Communities Railways,* Com(96)421.

Evans, A. & A. Morrisson (1997), Incorporating accident risk and disruption in economic models of public transport, *Journal of Transport Economics and Policy*, May.

Haldi, J. and D. Whitcomb (1967), Economies of scale in industrial plants, *Journal of Political Economy,* Vol. 75, No. 4.

Hansson, L. (1997), *The internalization of external effects in Swedish transport policy -- a comparison between road and rail traffic,* International Institute for Industrial Environmental Economics, Lund University.

Harris, R. (1977), Economies of traffic density, *Bell Journal of Economics,* Vol. 30, No. 4.

Hedström, R. (1996), *Samband mellan trafikbelastning -- banstandard -- underhållskostnader,* VTI-meddelande 806.

Hylén, B. (1995), *Access to rail infrastructure -- Regulation and pricing,* A state-of-the-art report from ten European countries, VTI, No. 42-1995.

Jansson, J.O., P. Andersson, P. Cardebring and T. Sonesson (1992), *Prissättning och finansiering av järnvägens persontransporttjänster,* TFB & VTI forskning/research No. 5.

Jansson, J.O. and G. Lindberg (1997), *Transport pricing principle,* D2 of PETS (Pricing European Transport Systems). Research project for the Commission of the European Communities: Transport -- DG VII.

Jansson, J.O. (1997*),* Theory and practice of transport infrastructure and public transport pricing*,* in: *Recent Developments in Transport Economics*, G. de Rus and C. Nash (eds.), Ashgate, Aldershot.

Johansson, J. (1997), *The economics of railway safety -- state of the art,* Department of Infrastructure and Planning, KTH, Stockholm.

Kopicki, R. and L.S. Thompson (1995), *Best methods of railway restructuring and privatisation*, CFS Discussion Paper Series, No. 111, The World Bank, August.

Ministry of Transport (1964), *Road pricing -- the economic and technical possibilities,* London (Smeed Report).

Mohring, H. (1972), Optimization and scale economies in urban bus transportation, *American Economic Review,* Vol. 62, No. 4.

Molinder, J. (1997), *Värdet av turtäthet,* Working paper, EKI, Linköping University.

Nilsson, J.-E. (1997), Efficiency implications of innovations in administering transport infrastructure, Department of Economics, Stockholm University.

Panzar, J. (1979), Equilibrium and welfare in unregulated airline markets, *American Economic Review, Papers and Proceedings*.

Pratten, C. (1971), *Economies of scale in manufacturing industry,* Cambridge University Press.

Scherer, F. and D. Ross (1990), *Industrial market structure and economic performance*, Houghton Mifflin Co., Boston.

SJ (1977), *Trafiksäkerheten 1996,* SJ Stab Trafiksäkerhet.

Swedish National Rail Administration (Banverket) (1997), Banverkets beräkningshandledning.

Swedish National Rail Administration (Banverket) (1997a), Översyn av banavgiftssystemet, Banverkets rapport enligt regeringsuppdrag 1997-05-29.

Walters, A. (1961), The theory and measurement of private and social cost of highway congestion, *Econometrica,* Vol. 29, No. 4.

Walters, A. (1968), *The economics of road user charges,* World Bank occasional paper No. 5, Washington.

NOTE BY THE CHAIRMAN

INTRODUCTION TO THE DISCUSSION

The ideas developed in this note refer to the French experience

Claude MARTINAND
Chairman
Réseau Ferré de France
Paris-La Défense

184

SUMMARY

Paris, March 1998

1. THE MAJOR ISSUES INVOLVED IN INFRASTRUCTURE PRICING

1.1. A vital signal for all the actors

The fact is that user charges constitute, without a shadow of a doubt, the main economic signal of the railway system, which will guide the short-term operating choices and medium- and long-term investment choices of the different actors concerned:

- The train operators using the network, and their clients;
- The infrastructure company (RFF) and its operator (SNCF);
- The territorial authorities, as the organising and cofinancing authorities of infrastructure and train services;
- The State as "regulator", the organising and supervising authority of the RFF and the SNCF;
- All the actors concerned by competition and co-operation between transport modes.

1.2. A factor for the lasting equilibrium of the railway system accounts

In addition, charging for infrastructure use and the methods of financing the network (operation), maintaining it in good order and developing it (investment) are indissolubly linked. The financial equilibrium of the infrastructure company has to be ensured from one year to the next and must be sustainable.

At any given moment, the infrastructure costs reflect at the same time the quality of the network, its state of maintenance and modernisation and the level of performance of the operator responsible for the maintenance and operation of the network (service quality and productivity).

The various costs of the rail network have to be covered by:

- Income from users (charges and fees);

- Public contributions (contribution to infrastructure costs, equipment subsidies and capital endowment);
- Possible contributions by indirect beneficiaries (transport and other levies paid into the FITTVN or the FARIF);
- Borrowings on the financial markets, which in fact merely transfer today's investment expenditure to tomorrow's users (or even taxpayers), at a not insignificant cost.

It is also worth listing the different benefits of general interest that may justify public contributions to transport infrastructure or services:

- Implementation of the public service obligations imposed for geographical or social reasons, which merits public contributions;
- Balanced regional development, implying equality of opportunity for all regions, thanks to adequate conditions of access at a reasonable cost to the user (and the community); but rail transport is not necessarily the only mode or the most appropriate one for this purpose;
- Targeted structuring of the regional, national or European space by means of investments orienting the localisation of activities and anticipating future traffic (cf. trans-European networks);
- Harmonization of infrastructure costs and intermodal competition (if there is no internalisation of the external or social costs) and incentive pricing to encourage the use of the transport mode which is safest, least polluting and most advantageous for society as a whole;
- Elimination of cost burdens inherited from previous imbalances in the infrastructure accounts (in the context of the application of Directive 91-440).

The nature and method of payment of these contributions are of considerable importance:

- Operation or investment;
- Infrastructure or services;
- Fixed *a priori* or *a posteriori* (subsidy to cover the deficit is to be avoided);
- Incentive or neutral.

The supervision of this "state aid" with respect to the rules laid down by European or national texts (competition law) is increasingly strict and not everything is possible any more.

It should be pointed out that the French rail reform of 1997 has not yet ensured the prospect of lasting financial equilibrium for the railway system. The completion of the financial aspects of the reform and the pricing system therefore cannot be dealt with independently from one another.

2. THE MAIN OBJECTIVES TO BE CONSIDERED

Charging for rail infrastructure use has been implemented in very different ways in different European countries. It has also been the subject of many analyses and studies, notably by the Commission (DG VII) and the Union Internationale des Chemins de fer (UIC), with a view to drafting recommendations or a directive on eventual harmonization, for example, for alternative and competing routes serving the ports.

In this context, it is appropriate to recall the main objectives that should be taken into account in the pricing system:

1. Cover all or part of the operating and maintenance costs of the rail network and reflect the level of service provided to the carrier. This coverage may be achieved globally or, on the contrary, for each section of the network, and this may be in identical or differentiated fashion;
2. Favour the best possible use of the rail network from the standpoints of the management of priorities in operation (routes/slots) and economic efficiency criteria (economic surplus, for example) and non-discrimination;
3. Contribute to the costs of developing the rail network through making investment self-financing;
4. Encourage the use of the rail transport in intermodal competition, because of the insufficient harmonization of the conditions of intermodal competition (external costs, social costs);
5. Contribute to the balanced regional development, through improving the accessibility of disadvantaged areas, for reasons of equity and solidarity.

These objectives are to some extent contradictory, it must be admitted, and any pricing policy will have to reflect the ranking of objectives established in the compromise accepted. Each option decided upon must be able to be evaluated from the standpoint of efficiency on the one hand and equity on the other.

In order to both reflect the costs more accurately and to permit a pricing policy differentiated according to the objectives decided upon, it is usual to adopt pricing formula with two or three terms and a segmentation of the network into sub-networks and basic homogeneous sections.

The tariff formula includes a reservation fee and an operating charge and, possibly, an access fee. Decree 97-446 effectively provided for these three terms. The operating charge seems well suited for translating objective 1 (cost coverage) and objectives 4 (harmonization of intermodal competition) and 5 (regional development), thanks to subsidies to avoid the total imputation of costs. The reservation fee seems well suited for translating objectives 2 (better use of the network) and 3 (contribution to the costs of developing the network), especially on the most heavily trafficked parts of the network.

Unless the pricing is determined section by section, the network can be segmented into sub-networks, in particular according to traffic volume and service quality. The provisional pricing system has opted for four sub-networks:

- R_0 : the most heavily trafficked lines, in particular around the major rail nodes (about 750 kilometres);
- R_1 : new high-speed lines (about 830 kilometres);
- R_2 : strongly developing intercity lines (about 4 630 kilometres);
- R_3 : other lines (the rest of the network).

Lastly, price differentiation according to time may be envisaged, either daily (peak hours) or annually (peak days), in particular on the most heavily trafficked network.

3. THE BASES FOR OPTIMAL PRICING

Theoretical studies by experts have led to optimal pricing proposals being made to decisionmakers. These are based on the different types of cost:

– Marginal cost (short-term), integrating the costs of congestion;
– Marginal social cost, integrating the externalities;
– Marginal social cost of development (long-term), integrating the cost of renewal investment;
– Total cost, but this cost depends on the conventions adopted for covering different expenditures on the operating account and, above all, the methods of financing investments and the corresponding amortization.

Economic theory concludes that optimal pricing should be based on the marginal social cost of development. However, because of the budget equilibrium constraints, a second-rank optimum, designed to cause the least possible distortion to the choices of the economic actors (Ramsey-Boiteux), may be adopted.

It should, in fact, be noted that our knowledge of the costs and how they vary as a function of different parameters is very fragmentary and inadequate.

To give an order of magnitude, it is considered that the present amount of 6 billion must be higher than the variable costs and the total cost for RFF can be estimated at over 23 billion for 1998 (excluding the present contribution to the infrastructure costs of 11.8 billion and excluding the financial costs of eliminating the debt of 134.2 billion).

If the contribution to the infrastructure costs is maintained at its present level (11.8 billion), an amount of 12 billion will balance the infrastructure account. This subsidy can then be justified by objectives 4 (harmonization of intermodal competition) and 5 (regional development).

As compared with the theoretical conclusions, it has to be said that the different European countries have opted for very different policies:

– Total cost in the United Kingdom and Germany, these countries preferring to subsidize the railways through public service agreements and investment subsidies;
– Marginal cost in Sweden, the equilibrium of the infrastructure account being ensured directly by the State.

There is, thus, very clearly much to debate and choices to be made between different options according to the different issues identified and the consequences of these choices.

4. THREE MAJOR CHOICES TO BE DEBATED

4.1. The overall level of user charges

- What overall level of charges should be aimed at, either in steps during a transition period, or right from 1999, with the State being prepared to make a contribution to the SNCF for its public service missions, permitting it to pay these charges without bringing its accounts into imbalance again?
- Would an overall total in the order of at least 12 billion, halfway between the present level and the result of total cost pricing, be the right compromise, taking account of the specificities of the French situation (extensive territory with a low and very uneven population density), regional development considerations implying a sufficiently dense network?

Such a level would significantly improve the self-financing capacity of the RFF and would reduce the public financing requirement for the development of the network. It would respect the rules stemming from the second phase of the Economic and Monetary Union Treaty and would avoid having to reclassify the RFF debt as public debt. It would ensure the financial equilibrium of the railway system while maintaining contributions to the RFF at their present level.

4.2. Price differentiation according to the network

- What price differentiation should be envisaged, combining the respective levels of the operating charges and reservation fees, according to the segmentation of the network?
- Is the present segmentation of the network appropriate? Should it be maintained or modified and, if so, how?
- Should the segmentation of the network be supplemented by a temporal differentiation, which would be introduced only on the saturated sections of the network?
- Can the reservation fee be devoted to objectives 2 (optimal utilisation of the network) and 3 (development costs) and the operating charge to objective 1 (cost coverage), the public contribution to operation being devoted essentially, if not exclusively, to objectives 4 (harmonization of competition) and 5 (regional development) and the public investment subsidies being devoted to objectives 3 (development

costs), 4 (harmonization of competition) and 5 (regional development)?

In short, for which part of the network should we try to achieve total cost coverage and on which part should we renounce this, provided that the corresponding public contributions are received?

4.3. Price differentiation according to the activity

- Depending on the reality of the intermodal competition, which varies greatly according to the activity and the market, should we envisage price differentiation according to the nature of the activity (freight, mainline services, regional and local services) and according to the origin or destination of the links (services to or from the ports, for example)? Is such an option to some extent the same as that of the "contributive capacity" advocated by the SNCF?
- Should the total cost pricing on certain parts of the network concern all train categories or only those that are determining factors for operation and for capacity increase requirements (the case of suburban trains in the Île-de-France, for example)?

5. PRACTICAL CHARGING METHODS

In addition to these three sets of major choices, practical charging methods also need to be clarified and discussed.

The charging should be stable (which argues in favour of introducing the final pricing system right away, with no transition stage), relatively simple and transparent, if we want it to produce all its incentive effects.

Any differentiation envisaged must be able to be effectively measured at reasonable cost if it is to be useful. Many refinements relating to the technical characteristics of the network or the trains are interesting in theory but rarely operational. To what extent should we take account of the train characteristics (power, length, axle weight, etc.) and the way in which it uses the infrastructure (speed, stops, etc.)?

- Charging is generally based on the train-kilometre. Should we add a charge connected with the time that a particular part of the infrastructure is used (a platform in a saturated station for example)?
- The access fee (fixed term) must not be discriminatory nor constitute a barrier to entry: at what level should this access fee be fixed? Should we envisage a possibility of choice (as in Germany), with a high access fee and lower variable charges or no charge for access and higher variable charges?
- Should we also introduce the possibility of negotiating the charges, as in the United Kingdom, according to the length of the contracts, or even income guarantees, or provide only for simple quantity discounts?
- Should we envisage reciprocal penalties and indemnities according to the respect of commitments of regularity, either by the infrastructure manager or by the train operators? (such a system is used only in the United Kingdom)?
- Should we implement articles 6 and 10 of Decree 97-446 relating to specific charges in the case of particular investments? Should we give up this idea or use it only when really justified, i.e. rarely, in order to avoid a kind of "balkanisation" of the pricing?

Many other questions are no doubt worthy of mention, at the risk of clouding the debate on the major issues. These will be raised at the appropriate time.

SUMMARY OF DISCUSSIONS

SUMMARY

INTRODUCTION

Against a background of historical decline in rail transport in Europe (except for high-speed passenger transport), railway reform has now entered the phase of implementation. After separating infrastructure from operations, on an accounting basis at least, governments now have to define the user charges that train operators will pay for the services provided by infrastructure managers. The very newness of the concept of infrastructure charging raises problems and countries have not all taken the same approach in the solutions adopted. At the same time, the development of European “freight corridors” as a means of facilitating international rail transport inevitably makes user charging an issue as soon as more than one network is involved. Underlying this issue are budgetary pressures in Europe that seek to reduce subsidies to rail networks.

What features must a user charging system have in order to reverse the historical decline of the railways: in other words, to make services more competitive, end deficits and reduce the reliance on subsidies?

The Round Table set out to answer these questions by determining:

- -- firstly, strategic objectives in the design of user charges;
- -- secondly, the basic principles for user charging systems;
- -- thirdly, practical points for the design of user charging systems.

1. STRATEGIC OBJECTIVES IN THE DESIGN OF USER CHARGES

It is the vertical separation of rail operations and infrastructure that brings the issue of user charges for railway infrastructure to the fore. The main objective in separating operations and infrastructure was to make the component costs of rail services transparent (use of track, equipment, manning), enabling inefficiencies to be better identified and controlled. Separating the accounts for operations and infrastructure is also prerequisite to introducing competition. Since competition encourages innovation and productivity gains, it could reduce the railways' reliance on subsidies and increase their modal share.

Once more than one undertaking has rights of access to infrastructure to run trains, a mechanism is required to ensure that fees and conditions for access are non-discriminatory. This can be promoted by setting rules for the structure (and/or level) of charges that can be levied. Among the many objectives that policymakers may seek to achieve through such rules are:

- Cost recovery (possibly including external costs) from users;
- Better utilisation of infrastructure capacity;
- Guidance for investment choices;
- An increase in the market share for railways as opposed to other modes;
- A reduction in the costs of rail transport and an increase in productivity;
- An improvement in the quality of rail services.

Moreover, the problem of infrastructure charges is linked to that of infrastructure funding. A balance has to be struck between these imperatives and different countries address this in different ways.

To date, half the networks in ECMT countries have defined principles for the formulation of charges, at least for freight transport. However, at present, the level and structure of these charges vary widely from country to country. This is due in part to pricing polices which involve trade-offs between geographical, historical and societal considerations. For example, countries with remote or disadvantaged areas tend to impose low charges for serving them. Moreover, it is difficult to speak of any move towards uniformity, given the extent to which pricing practices are shaped by divergent railway reforms. In the United Kingdom, for example, Railtrack -- the infrastructure manager -- has been privatised, and user charges must be such as to generate returns on the company's assets.

It is readily seen, then, that pricing rules are tied in with how railway reform has been organised -- given that the relations between the players involved in rail services differ enormously from one country to another -- and with the strategic objectives that have been assigned thereto. These objectives are many and sometimes conflicting, and one might well question whether user charges for railway infrastructure are necessarily the best means of pursuing policy goals. Another result of this multiplicity of objectives is that any efforts at European harmonization will necessarily be very difficult. Setting multiple goals for the charging system is likely to make it extremely complicated, when what is needed to attract new rail operators is a straightforward, transparent charging system. Given the variety of national pricing arrangements, the Round Table sought to highlight the essential elements that economic theory suggests should be the basis of all rail infrastructure charging systems.

In a market economy, pricing serves to guide decisions in such a way as to encourage the more efficient use of available resources. Prices are a management tool which gives the system new possibilities for development. This is how the introduction of user charges for railway infrastructure must be seen -- as a powerful instrument for modernising the railways and improving their performance, i.e. for making them better able to meet competition from other modes of transport. In order to bolster the railways' share of aggregate transport, it is necessary to tackle costs, network quality and the productivity of all railway companies, whether they provide infrastructure or operate services. Accordingly, a system of user charges should serve to optimise network management, i.e. to improve capacity utilisation, orient investment decisions and encourage productivity gains throughout, thereby making it possible to reduce recourse to government funding.

There is a direct and very rational link between access charges and the economic value of infrastructure: pricing and capacity management are related elements. If infrastructure access is not charged at the real cost of making it available, service operators will be inclined to overuse it, because they will not have to pay the actual costs of providing it. Prices that do not reflect the value of the resource used may lead to excessive demand for, or inefficient use of, that resource. That is generally the situation which has prevailed until now. Public funding in the form of subsidies made up infrastructure account deficits, where such an account existed. With infrastructure use failing to generate adequate returns, expansion was impossible without government subsidies. These are an unreliable resource as always at the mercy of budget restrictions.

The primary economic objective of user charging should thus be to encourage rational use of infrastructure. With appropriate charges, i.e. charges that reflect the costs of making the infrastructure available, the economic utility of the infrastructure will become clear from the amount of use made of it. The revenues generated from user charges will give a signal to infrastructure managers that will tell them if infrastructure development will be efficient. The network will thus be managed by demand instead of by subsidies, which are often poorly planned and arbitrary. The principles of user charging for infrastructure will largely determine investment policy. This should clarify risks that were often not apparent in integrated undertakings whose goals were often far too broad-ranging, complex and conflicting.

Clear, unambiguous objectives have to be set for the railway system if it is to be rescued from its current plight. From the above, what might be called "cost regulation", therefore has to be a priority objective. This would put an end to following too many conflicting objectives. Also, the goal of cost regulation is perfectly consistent with the primary objective of railway restructuring, which is to introduce transparency and competition into the railway system: in a market economy, firms compete on the basis of their respective production costs, which presupposes that they are in full control of their internal pricing structures, hence the notion of transparency.

What features must a pricing system have in order to achieve this?

2. BASIC PRINCIPLES

The fundamental principle for incorporation in a system of user charges is to relate the charge for running one additional train to the additional costs entailed. This is the marginal-cost pricing principle. Such marginal costs are incurred through extra wear and tear on tracks, signalling operations, the administrative costs generated by additional trains and, where applicable, the electric power consumed. These particular marginal costs are designated as "short-run".

If a network is congested, as happens when capacity is inadequate, a congestion charge must be added to shift demand to less congested periods, or to finance additional capacity. One might add that automatically increasing capacity to meet demand does not necessarily ensure optimal resource allocation: discouraging use when infrastructure is congested can be perfectly justifiable.

Assuming that capacity is to be increased, the additional costs incurred are long-run marginal costs. These long-run marginal costs should be passed on to train operators, to encourage them to improve their operations, i.e. how they manage their requests to run trains. Once they have to "pay" the full costs of running trains out of their own pockets, train operators will be encouraged to opt for the least expensive train paths. It should be noted that congestion costs become very real when infrastructure is close to capacity, particularly since capacity is never optimal -- it can only be extended in steps, not continuously. It is therefore important on congested lines to allocate scarce paths by making operators pay the price.

One might conclude from this that, in order to optimise user charges for railway infrastructure, i.e. to promote rational use thereof, pricing should be based on long-run marginal costs. However, it should be noted that it is rather difficult to evaluate long-run costs when several different operators are competing for paths. Where this is the case, the costs of the additional capacity needed by one train operator, which translate the long-run marginal costs into operational terms, are determined by a series of approximations. Marginal costs in general are difficult to measure and we have to make do with reasonable approximations.

Experience has shown that a pricing system based on marginal costs cannot enable the infrastructure provider to break even if the network has large capacity surpluses. Networks are characterised by economies of density and scale -- short-run marginal costs decline as the network grows -- which means that marginal cost pricing leaves some costs uncovered. In such cases where marginal costs are lower than average costs, they should be supplemented by fixed charges. Indeed, the Round Table concluded that this was essential for a pricing system that does not lead to infrastructure account deficits. After all, it makes more sense to finance contracted services than to finance a deficit. Thus, in order to recoup costs -- a priority for the Round Table -- in addition to the variable element corresponding to short-run marginal costs, charges should include a fixed element. This would give us a two-part charge -- short-run marginal costs plus a fixed element -- that would have the advantage of ease of use, a feature that should not be overlooked, in the view of the Round Table. However, there should also be incentives for infrastructure managers to improve efficiency and to lower user charges in relative terms.

Given a choice between financing the deficit of a train operator or that of an infrastructure provider, the Round Table felt it was better to assist the operator of train services. In fact, the Round Table went one step further: if appropriate decisions are to be made throughout the entire chain of railway services, it would be better to subsidise the end user and institute a pricing system that reflects true costs. Contrary to what one might think, it is because railways have not been subjected to the rules of a market economy that Europe's railway undertakings find themselves in their current situation.

One consideration is that such pricing principles, with a fixed element to ensure break-even, would penalise remote areas with little traffic. In this case, to maintain services, subsidies could always be maintained. However, even then it would be preferable that funding should go to subsidise operators rather than infrastructure deficits, i.e. to subsidise the end-user. Furthermore, focusing on real costs would provide a sounder basis for decisions on whether or not to maintain loss-making lines.

To take account of the externalities (such as noise) generated when an additional train is run, pricing could be based on long-run marginal social costs, by factoring in social costs which are not directly internalised. However, in view of the difficulties in evaluating external costs, the Round Table took the view that the externalities problem confused the issue of infrastructure user charging and served only to delay the institution of a vitally urgent pricing system. Ideally, a pricing system based on long-run marginal social costs would be adopted for all modes. This would avoid the distortion of competition between modes that makes it difficult to solve the problems of the railway undertakings.

Lastly, it should be noted that any theory on better pricing along the above lines, assumes that charge-payers will behave rationally and avoid taking expensive routes. In this connection, some Round Table experts were of the opinion that established carriers did not have a long tradition of rational economic behaviour.

In conclusion, the following key principles can be highlighted.

- Infrastructure pricing and investment should be linked, and this is where congestion plays a key role; infrastructure charges should preferably be demand-based;
- Fixed costs should be covered. Any subsidies required in the interest of social welfare to offset resulting price rises, should be targetted as closely as possible to the end user;
- The use of a two-part tariff presents an opportunity to avoid excluding traffic which cannot afford to pay a substantial share of fixed charges, with a minimum tariff corresponding to short-run marginal costs;
- Pricing should be transparent, which requires simplicity and a good information system.

3. PRACTICAL DETAILS

To preclude any risk of misinterpretation, user charges for railway infrastructure should be set in a manner that is simple, transparent, non-discriminatory and stable. Non-discrimination was discussed at length at the Round Table. Some experts thought it important to offer more favourable terms to new railway undertakings than to well-established carriers whose size already guaranteed them preferential terms. Setting low access charges and restricting reductions for large quantities would encourage competition, which is one of the basic objectives of reform. However, the new railway undertakings appearing in Europe belong to large groups that have substantial capital and therefore need not be given preferential treatment. The demand for paths is really not a “spot” market, where initiatives would be taken with no consideration for their implications.

A differentiated tariff is desirable so that those who want limited access rights will not have to pay a large proportion of common costs. It is important not to kill off potential demand. With that in mind, it is essential to set up appropriate user-friendly information systems taking advantage of leading-edge technologies, such as Internet.

Negotiating procedures should also be instituted alongside posted tariffs in order to respond with some flexibility to market imperatives: congestion costs -- i.e. the costs of expanding capacity -- can vary widely, depending on circumstances, and theoretical pricing models should be applied with some degree of latitude.

It should also be noted that the possibility of cross-subsidisation, which benefits large enterprises -- the only ones able to spread fixed costs over different segments of the market -- should be eliminated. Cross subsidies do not make for transparency, which is a prerequisite for the success of railway reform.

It may be worthwhile, although complex, to impose penalties on operators responsible for disrupting the system by, for example, causing delays. Thus both parties would be constrained to meet their commitments.

In Europe, the harmonization of user charges for infrastructure is hampered by the disparity of the reforms under way -- but progress can be made by establishing priorities: harmonized pricing structures could be adopted for international freight corridors and for the high-speed network. This would yield single pricing formulas conducive to the development of international rail transport.

For some of the Round Table's experts, ideally, infrastructure funding in Europe should be harmonized, which would make it much easier to devise a uniform user charging system. Failing this, a viable short-term alternative could be to cover infrastructure costs to the same extent on preferred routes. Consideration could also be given to setting up an international system for infrastructure managers that would offset current disparities in infrastructure cost coverage. In the longer term, an increase in road transport taxes to internalise externalities -- in the form of a sophisticated road pricing system -- would pave the way for uniform coverage of infrastructure costs in the transport sector. In any event, it is certain that international co-ordination will be needed to achieve an efficient system.

Train paths can be allocated in a variety of ways. It is preferable to institute a dynamic process which makes each of the companies concerned aware that any given path entails a certain cost. Insofar as the infrastructure provider is in a monopoly position, its decisions must be subject to appeal to an arbitrator assigned to moderate the viewpoints of the parties in question. For some experts, it is vital to introduce a specific structure for this purpose, which could also advise governments on railway organisation and on investment decisions in particular. Clearly, infrastructure managers, which are in a monopoly situation, might be tempted to earn an economic rent from train operators instead of investing in track renovation and capacity.

Moreover, it would be aberrant and contrary to the spirit of railway liberalisation if the traditional carrier enjoyed preferential rights over certain paths. User charges and train path allocations should not favour one enterprise at the expense of another. The Round Table pointed out that the task of allocating paths was complicated by the fact that traditional carriers operate suburban trains, which often have preferential access at peak hours and are heavily subsidised. Where this is the case, some compromise on policy will have to be reached.

CONCLUSION

The potential opening up of rail networks to third parties marks a decisive phase in railway reform in Europe. User charging is the *quid pro quo*, in economic terms, for the possibility of having more than one railway operator providing services. As things stand in Europe, there is very little effective competition between undertakings operating services on the same network. User charging systems must be designed to allow effective competition or, at least, rational use of infrastructure.

Pricing must not disguise the fact that reducing the costs of the railways is a key issue. In order to do so, the actual costs of running a train must be transparent. This is the only way in which all the players in the railway system will be able to take efficient decisions.

Relating the running of a train to the costs of operating it means taking marginal costs as the basis for pricing, ideally, long-term marginal social costs, including the costs of expanding infrastructure where it is congested and externalities. To ensure, at the same time, that infrastructure costs are fully covered, a two- or three-component pricing structure, which includes a fixed charge, is recommended. Within this fixed element, care should be taken to ensure that potential new entrants seeking low-volume operations are not excluded through a high access charge. The Round Table felt that it was preferable to cover train operators' losses, i.e. to subsidise end-users, rather than the deficits of infrastructure operators.

In Europe, having a range of unco-ordinated charging systems complicates the issue for international transport enormously. Given that this is the case, it would be preferable to harmonize the structure of the different charging systems and, at a later stage, to standardize the extent of coverage of infrastructure charges. This would avoid charging different rates for international transport on different routes.

To sum up, the methods recommended by the Round Table are based on marginal costs. This is the ideal theory and it should be stressed that departure from it will result in deficits and waste of resources for the community. Lastly, to those who would object that the concept of marginal costs which underpins the Round Table's proposal is a difficult one to apply, it can be responded that it is possible to make reasonable estimates which provide valid approximations.

LIST OF PARTICIPANTS

Monsieur Claude MARTINAND **Chairman**
Président
Réseau Ferré de France
Tour Pascal A
F-92045 PARIS-LA DEFENSE

Prof. Dr. Gerd ABERLE **Rapporteur**
Justus-Liebig-Universität Giessen
Lehrstuhl VWL I
Licher Strasse 62
D-35394 GIESSEN

Monsieur Luc BAUMSTARK **Co-rapporteur**
Commissariat Général du Plan
5 rue Casimir Perier
F-75007 PARIS

Monsieur Alain BONNAFOUS **Co-rapporteur**
Directeur de la MRASH
Laboratoire d'Économie des Transports (LET)
MRASH
14 avenue Berthelot
F-69363 LYON CEDEX 07

Mr. John DODGSON **Rapporteur**
Senior Consultant
National Economic Research Associates - NERA
15 Stratford Place
GB- LONDON W1N 9AFN

Prof Jan Owen JANSSON **Rapporteur**
Professor of Transport Economics
Linköping Universiteit
Linköping Högskola
S-581 83 LINKÖPING

Monsieur François BEFAHY
General Manager - “Voyageurs National”
SNCB
85 rue de France
B-1060 BRUXELLES

Mr. Magnus BØEN
Permanent Adviser
Ministry of Transport and Communications
P O Box 8010 Dep
Akersgt 59
N-0030 OSLO 1

Mr. K. CELINSKI
Director of Department of Railways
Ministry of Transport and Maritime Economy
6/4 Chalubinskiego str
PL-00 928 WARSAW

Monsieur Gunnar GUSTAFSSON
Directeur du Département Infrastructure
Union Internationale des Chemins de Fer
16 rue Jean Rey
75015 PARIS

Mme Dagman HAASE
Director Marketing and Sales
Deutsche Bahn AG
Stephensonstrasse 1
D-60326 FRANKFURT (Main)

Mr Jan HODGSON
Communautés Européennes
DG VII B2, BU31, 7ème étage
200 rue de la Loi
B-1049 BRUXELLES

Mr. Bertil HYLEN
Railway University
Swedish National Road and Transport Research Institute
S-581 95 LINKÖPING

Monsieur Jean-François JANIN **Observer**
Sous-Directeur - DTT
Ministère de l'Équipement, des Transports et du Logement
Grande Arche de la Défense Paroi Sud
F-92055 LA DEFENSE CEDEX

M. le Professeur Andres LOPEZ PITA
Universidad Politecnica de Cataluna
Catedratico de Ferrocarriles
Gran Capitan s/n Modulo B1
E-08034 BARCELONE

Mr Harald MINKEN
Chief Research Economist
Institute of Transport Economics (TOI)
Postboks 6110 Etterstad
N-0602 OSLO

Prof. Christopher NASH
University of Leeds
Institute for Transport Studies
GB- LEEDS LS2 9JT

Mr Jan-Eric NILSSON
Dalarna University
Centre for Research on Transportation
and Society - CTS
S-781 88 BORLANGE

Monsieur David NIVEN REED
Chargé de Mission
Union Internationale des Chemins de Fer
16 rue Jean Rey
75015 PARIS

Mr Marco PONTI
Professor of Transport Economics
TRT Trasporti e Territorio SRL
Piazza Arcole 4
I-20143 MILAN

Mr David REAMS
Office of the Rail Regulator
1 Waterhouse Square
138-142 Holborn
GB- LONDON EC1N 2ST

M. le Professeur Robert RIVIER
École Polytechnique Fédérale de Lausanne
Département de Génie Civil
ITEP - TSO
CH-1015 LAUSANNE

Mr. Nicolas ROODT
Ministry of Transport, Publics Works and Water Management
Department of Infrastructure Policy - IB/B
PO Box 29 901
NL- 2500 EX THE HAGUE

Mr. Gottfried SCHMITT
Nahverkehrsgesellschaft Baden-Württemberg
Wilhemsplatz 11
D-70182 STUTTGART

Mr. Robin TUCKER
27 Old Deer Park Gardens
GB-RICHMOND TW9 2TN (Surrey)

Mrs. Caroline VAN DER MAAS
Ministry of Transport, Publics Works and Water Management
Transport Research Centre AVV
PO Box 1031
NL-3001 BA ROTTERDAM

Professeur José Manuel VIEGAS
CESUR - IST
Av. Rovisco Pais
P-1096 LISBONNE CODEX

Mr. Jerzy WRONKA
Research Institute for Transport Economics
Korzeniowskiego 1
PL-70 211 SZCZECIN

ECMT SECRETARIAT

M. Gerhard AURBACH - Secretary-General

Mr. Jack SHORT, Deputy Secretary-General

ECONOMIC RESEARCH, STATISTICS AND DOCUMENTATION DIVISION

Mr. Alain RATHERY, Head of Division

Mr. Michel VIOLLAND, Administrator

Ms Françoise ROULLET, Assistant

Mrs Julie PAILLIEZ, Assistant

TRANSPORT POLICY DIVISION

Mr. Stephen PERKINS - Principal Administrator

ALSO AVAILABLE

Express Delivery Services. Series ECMT - Round Table 101 (1996)
(75 96 04 1) ISBN 92-821-1214-4 France FF110 Other Countries FF 145 US$28 DM42

Changing Daily Urban Mobility: Less or Differently?. Series ECMT - Round Table 102 (1996)
(75 96 06 1) ISBN 92-821-1216-0 France FF260 £34 US$50 DM76

The Separation of Operations from Infrastructure in the Provision of Railway Services. Series ECMT - Round Table 103 (1997)
(75 97 02 1P) ISBN 92-821-1221-7 France FF295 £38 US$58 DM86

New Trends in Logistics in Europe - Round Table 104 (1997)
(75 97 05 1 P) ISBN 92-821-1224-1 France FF215 £28 $US42 DM63

Infrastructure-Induced Mobility. Series ECMT - Round Table 105 (1998)
(75 98 07 1 P) ISBN 92-821-1232-2 France FF400 £40 $US67 DM119

Intercity Transport markets in Countries in Transition. Series ECMT - Round Table 106 (1998)
(75 98 10 1 P) ISBN 92-821-1235-7 France FF400 £41 $US66 DM119

Rail Restructuring in Europe (1998)
(75 98 08 1 P) ISBN 92-821-1233-0 France FF180 £18 $US30 DM54

Prices charged at the OECD Bookshop.

The OECD CATALOGUE OF PUBLICATIONS and supplements will be sent free of charge on request addressed either to OECD Publications Service, or to the OECD Distributor in your country.