User Charges for Railway Infrastructure: French experience
Luc Baumstark, Alain Bonnafous

To cite this version:

HAL Id: halshs-00319740
https://halshs.archives-ouvertes.fr/halshs-00319740
Submitted on 9 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
With the authors and the publisher of the copyright agreement. The present version in PDF is under the copyright of CEMT / ECMT (http://www.internationaltransportforum.org/home.html).

This document is an extract of a report of ECMT. User Charges for Railway Infrastructure. Round Tables: No. 107, 26-27 March 1998, Paris (France).
(Complete Edition - ISBN 9282112403)
FRANCE

Luc BAUMSTARK
Chargé de Mission
Commissariat Général du Plan
Paris

Alain BONNAFOUS
Professeur
Laboratoire d'Économie des Transports
Lyons
France
SUMMARY

1. INTRODUCTION ..................................................................................... 51

2. INFRASTRUCTURE PRICING, KEYSTONE OF
   INSTITUTIONAL INNOVATION........................................................... 53
   2.1. Network opening, redistribution of roles and pricing system .......... 53
   2.2. Degree of vertical separation in the rail sector and the
       question of access charges and user fees ........................................... 55

3. INFRASTRUCTURE PRICING AND MAXIMUM EXTRACTION
   OF THE AVAILABLE SURPLUS ........................................................... 59
   3.1. The concept of marginal cost and the optimisation
       of investment ..................................................................................... 60
   3.2. The supply-demand equation or how to define an
       optimal quality of service ............................................................... 64
   3.3. Allowing for the budgetary constraint.............................................. 69

4. APPLICATION OF PRINCIPLES AND NETWORK
   CONSTRAINTS........................................................................................ 76
   4.1. The British experience: making a surplus ...................................... 76
   4.2. The German experience: covering the costs ...................................... 79
   4.3. The French example: a transitional phase ...................................... 85
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.

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
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.
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. 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, 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, 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 $I$ and the level of production $q$:

$$C(I,q) = a(I) + b(I,q)$$

Thus it is possible to distinguish between two marginal costs:

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

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

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

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

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 $C(I)$. 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.

“If 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?“

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”.

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. 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. 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.

Let us suppose that all the users 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 \( P(Q_i, g) \), 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, \( \psi(k) \).
Taking into consideration several time periods $t$, we write $P_i(Q_{it}, g_t)$, 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, $P_i(Q_{it}, g_t)$ 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 $D_i$ 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:

$$P_{it} - Q_{it}V_{it}D_i(Q_{it}, ..., Q_{nt}, k)$$

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:

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

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

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

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:
The relation for the investment is written as follows:

\[
\frac{\partial S}{\partial Q_{it}} = P_{it} - V_{it} D_t (Q_{1t}, Q_{2t}, \ldots, Q_{nt}, k) - \sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\partial D_t}{\partial Q_{it}}
\]

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

\[
\begin{align*}
\frac{\partial S}{\partial k} = -\sum_{i=1}^{T} \sum_{j=1}^{n} V_{jt} Q_{it} \frac{\partial D_t}{\partial k} - \Psi'(k) \quad (2) \\
\end{align*}
\]

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:

\[
\sum_{j=1}^{n} Q_{jt} V_{jt} \frac{\partial D_t}{\partial Q_{it}} \quad (3)
\]

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 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.” 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 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.”

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. Thus the balanced budget constraint is aimed at an efficiency deficit which goes far beyond the traditional criticisms levied 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\textsuperscript{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\textsuperscript{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, 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\textsuperscript{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\textsuperscript{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\textsuperscript{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, 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 market economy from the standard model. 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 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

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. 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


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 are based on a cost allocation study that uses the concepts of avoidable and additional 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\(^1\).
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

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


| 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

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 R0) and to a lesser extent on the busy high-speed lines (R1). On the other hand, on the low-density, high-speed network and on the main-line network (R2) the fees are very low, while on the rest of the network (R3) 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)

<table>
<thead>
<tr>
<th>Sub-network</th>
<th>$R_0$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>11 000</td>
<td>11 000</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>RC (peak)</td>
<td>100</td>
<td>18</td>
<td>0.85</td>
<td>0</td>
</tr>
<tr>
<td>RC (normal)</td>
<td>44</td>
<td>6</td>
<td>0.85</td>
<td>0</td>
</tr>
<tr>
<td>RC (slack)</td>
<td>20</td>
<td>4</td>
<td>0.85</td>
<td>0</td>
</tr>
<tr>
<td>TC</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
</tbody>
</table>

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.


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.


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.


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.


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.


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

BIBLIOGRAPHY


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.


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


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


