VATT-KESKUSTELUALOITTEITA
VATT-DISCUSSION PAPERS

245

HELSINKI WORKSHOP ON INFRASTRUCTURE CHARGING ON RAILWAYS

31 JULY - 1 AUGUST, 2000

Chris Nash
Esko Niskanen (eds.)

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Helsinki 2000

ISBN 951-561-351-5
ISSN 0788-5016

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Hämeentie 3, 00530 Helsinki, Finland
Email: cnash@its.leeds.ac.uk
esko.niskanen@vatt.fi

Yliopistopaino
Helsinki, December 2000

Preface

On behalf of the Government Institute for Economic Research (VATT), I thank the authors of the papers in this Conference Proceedings and thank all the participants of the Helsinki Workshop on Rail Infrastructure Pricing. Your contributions made the workshop successful and made these proceedings possible. Many authors kindly prepared revised versions of the papers afterwards. I would like to express my particular gratitude to Chris Nash for scientific co-operation, from the early planning of the workshop to the finalising of this report. Also, my special thanks go to Kaisa Kotakorpi and Anita Niskanen for skilled staff support, and to Anita Niskanen for work in editing.

Esko Niskanen

Research Director

CHRIS NASH – ESKO NISKANEN (eds.): HELSINKI WORKSHOP ON INFRASTRUCTURE CHARGING ON RAILWAYS 31 JULY - 1 AUGUST, 2000. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research, 2000, (C, ISSN 0788-5016, No 245). ISBN 951-561-351-5.

Abstract: Whilst there is a long tradition of economic analysis of charging for the use of road infrastructure, the situation relating to rail is quite different. In most cases until recently railways have been integrated organisations with the same body responsible for infrastructure and operations. Rail infrastructure charging, if practised at all, has therefore been largely an internal transfer within the organisation. The principal exception to this has been the presence of access rights by one company over infrastructure owned by another in the case of North America. But the dominance of integrated national rail companies in Europe, and the fact that international traffic was operated by cooperation between the national railways rather than by one railway operating over the tracks of another, made the issue of no relevance here. The change in this situation started with the separation of infrastructure from operations in Sweden, which was undertaken with the aim of putting rail infrastructure on an equivalent footing to road. Later on the European Commission began pushing for separation of infrastructure from operations in order to encourage new entry and to promote competition between different rail companies.

Key words: infrastructure charging, infrastructure financing, competition, market structure, railways

CHRIS NASH – ESKO NISKANEN (eds.): HELSINKI WORKSHOP ON INFRASTRUCTURE CHARGING ON RAILWAYS 31 JULY - 1 AUGUST, 2000. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research, 2000, (C, ISSN 0788-5016, No 245). ISBN 951-561-351-5.

Tiivistelmä: Tieliikenneinfrastruktuurin käytön hinnoittelua on taloustieteessä analysoitu jo pitkään. Rautatieliikenteen osalta tilanne on ollut kovin erilainen. Useimmissa maissa on viime aikoihin asti yksi ja sama organisaatio vastannut sekä infrastruktuurin tuottamisesta ja ylläpidosta että sillä tapahtuvan liikenteen järjestämisestä. Rautatieinfrastruktuurin hinnoittelussa, jos sitä on toteutettu ollenkaan, on paljolti ollut kyse organisaation sisäisistä maksuista. Merkittävä poikkeus on ollut yhden yhtiön oikeus käyttää toisen yhtiön omistamaa infrastruktuuria Pohjois-Amerikassa. Euroopassa kansallisten integroitujen rautatieyhtiöiden valta-asema ja kansainvälisen liikenteen hoitaminen näiden yhtiöiden välisenä yhteistyönä vaikutti siihen, että infrastruktuurin hinnoittelu ei ollut ajankohtainen kysymys täällä. Tilanne muuttui kun infrastruktuurin ylläpito erotettiin liikenteen hoidosta ensin Ruotsissa ja myöhemmin muulla. Muutoksen alkuperäisenä tarkoituksena oli saattaa rautatie- ja tieliikenne kilpailullisesti yhdenvertaiseen asemaan. Myöhemmin on Euroopan komissio ajanut rautatieinfrastruktuurin ja liikenteen hoidon erottamista tarkoituksena edistää alalle tuloa ja rautatieyhtiöiden keskinäistä kilpailua.

Asiasanat: infrastruktuurin hinnoittelu, infrastruktuurin rahoitus, kilpailu, markkinarakenne, rautatiet

Contents

Seminar on Infrastructure Charging on Railways – an Introduction

Chris Nash[*] *and Esko Niskanen*[**]

Whilst there is a long tradition of economic analysis of charging for the use of road infrastructure, the situation relating to rail is quite different. In most cases until recently railways have been integrated organisations with the same body responsible for infrastructure and operations. Rail infrastructure charging, if practised at all, has therefore been largely an internal transfer within the organisation. The principal exception to this has been the presence of access rights by one company over infrastructure owned by another in the case of North America, and this situation has indeed led to developments in economic theory by some of the foremost economists in the field, such as Baumol and Willig. But the dominance of integrated national rail companies in Europe, and the fact that international traffic was operated by cooperation between the national railways rather than by one railway operating over the tracks of another, made the issue of no relevance here.

The change in this situation started with the separation of infrastructure from operations in Sweden, which was undertaken with the aim of putting rail infrastructure on an equivalent footing to road. Thus from the first rail infrastructure pricing in Sweden was related to short run marginal cost, with balancing charges per vehicle in service equivalent to those found in road transport. But when the European Commission began pushing for separation of infrastructure from operations in order to encourage new entry and to promote competition between different rail companies, the issues involved became more complex, as demonstrated by the papers in this workshop.

In the opening paper, *Chris Nash* explains how the characteristics of rail transport, namely economies of scale, density and scope, lead to a trade-off between alternative objectives of infrastructure pricing. Prices cannot simultaneously promote efficient use of the infrastructure, efficient development and investment, meet budget constraints, promote competition and achieve harmonisation between modes. British experience to date has been a combination of simple two part tariffs for franchises rights enshrined within passenger franchises, and individual negotiations. The current review of charges by the Regulator is seeing a move towards a somewhat more complex tariff, in order to

[*] Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK,
email: cnash@its.leeds.ac.uk

[**] Government Institute for Economic Research (VATT), P.O. Box 269, FIN-00531 Helsinki,
email: esko.niskanen@vatt.fi

reflect marginal social cost more accurately, but with less need for individual negotiation.

Gunnar Gustafsson and A. Knibbe emphasise one problem caused by the diversity of charging structures and levels throughout Europe – the missed opportunity for international traffic, and particularly for freight. Although infrastructure charges are on average not a large part of total cost of rail freight transport, the variety of charging levels and systems, and the resulting complexity and lack of transparency, are argued to be a major handicap to the growth of international rail traffic.

Following a presentation by *Tom Howes* on the current state of EC policy towards rail infrastructure pricing, and an explanation of the so-called railway package agreed by the Council of Ministers in 1999, a lively debate ensued. This centred on whether the provisions for deviations from marginal cost pricing so weakened the proposed Directive as to make it worthless, or whether the requirement that such deviations should not lead to significant loss of traffic could be enforced to ensure that the main purpose of the Directive was still fulfilled.

Amihai Glazer introduced a different argument, that individuals and firms might prefer simple price structures to complex ones because complex price structures brought with them the likelihood that mistakes would be made. He presented some evidence that complex price structures were an inferior good, and that as countries become richer so price structures become simpler.

A. Bassanini and Jérôme Pouyet then examined the theoretical issues arising when infrastructure is nationally owned and vertically separated operators run international services. They show that two problems arise; one is the familiar double marginalisation effect, whereby any mark-up applied by one of the infrastructure owners is treated as part of marginal cost by the other, and the other is the constituency effect, whereby even the government of one country will ignore the effect on welfare of the other of their decisions. The result is sub-optimal pricing of the rail system as a whole, with each country trying to free ride on the other, and providing sub-optimal subsidies to infrastructure as a result.

As commented earlier, infrastructure pricing is only really an issue when there is at least a degree of separation between infrastructure and operations. *Pedro Cantos Sánchez* estimated translog cost functions for European railways in order to test whether such separation was efficient. He found that whilst freight and passenger operating costs were independent, freight and infrastructure costs were complementary whilst passenger and infrastructure costs were substitutes. His conclusion is that the inefficiencies from separation of infrastructure and operations must be fully evaluated in any assessment of the desirability of such a policy.

Bernard Caillaud and Jean Tirole examined the interrelationship between open access and the financing of investment in new or enhanced infrastructure, in circumstances in which the incumbent possesses private information on profitability. They show that the award of an exclusive franchise makes it more likely that the franchisee will fund investment, but more disturbingly they also show that the incentive to award exclusivity to fund investment is higher the stronger the demand for the service and therefore the greater the case for competition.

Stephen Gibson presented the results of a study of marginal costs of congestion as a result of the introduction of an additional train into the timetable. Whilst in Britain an operator is currently charged, via the performance regime, for the delays its trains impose directly on other trains, there is a further effect in that the presence of additional trains causes the repercussions of delays caused by other trains to be more severe. He showed results demonstrating that these delays can be very substantial; he also showed that they are greatly reduced if additional flexibility is introduced into the precise timing of the additional trains. Railtrack is proposing that charges to cover these costs should be included in the variable part of the two part tariff paid by British passenger franchisees.

The following two papers were triggered by consideration of the situation in two of the major countries in Europe, Germany and France. *Heike Link* showed that the existing infrastructure charging regime was a very significant constraint on new entry, in a country where the existence of a large number of local railways made new entry a realistic possibility. *Emile Quinet* introduced a degree of pessimism about what can be achieved by the reform of infrastructure charging regimes by showing the impossibility of achieving optimality in path allocation by pricing alone. The reason is that the optimal allocation must already be known in order to calculate the optimal price. Auctioning of access rights was an attractive way of solving the allocation and pricing problems simultaneously, but there were many practical problems. Consequently it may be that the traditional approach of imposing priority rules is the best that can be done in practice, but these rules need to pay more attention to the social value of the train rather than follow simple commercial priorities.

The workshop closed with further experiences by a range of parts of the world. *Per-Ove Hesselborn* and *Tiina Idström* presented experiences from Sweden and Finland respectively, and results showed that for both countries short run marginal cost pricing of rail infrastructure maintenance costs would only recover a very small proportion of total maintenance costs. *Katalin Tanczos and Gyula Farkas* described the relatively complex emerging structure of charges for infrastructure use being implemented in Hungary as part of the rail reform process. Finally, *Javier Campos* presented South American experience of rail franchising and considered the conditions necessary for its successful achievement. He concluded that it was very important to think through all the relevant issues, including access conditions and infrastructure charges, before embarking on the franchising process in order to minimise the uncertainties involved.

This wide range of papers provides a host of valuable theoretical and practical lessons. But it fails to come up with a single optimal policy for rail infrastructure charges. That is not surprising. The relative merits of the different approaches depend on a whole host of factors, including whether the regime involves exclusive franchising or provides for open access, the availability of alternative sources of investment funds and whether in practice there is a powerful incumbent or a range of operators. If funding is unlimited, then the best approach will be one based on short run marginal cost pricing, but this will not solve the allocation problem unless scarce slots can be auctioned off. If it is necessary to raise more funding than this, there is a trade-off between relatively sophisticated forms of second best pricing which minimise distortions but reduce transparency and hamper competition, and simpler approaches which are however more distorting. Clearly where exclusive franchises are in place, the former policies will be favoured, whereas a country which places a lot of emphasis on open access may favour the latter approach.

The result that the optimal approach to rail infrastructure charging varies so strongly with circumstances means that the prospects for the degree of harmonisation throughout Europe seen as necessary to promote international rail traffic may seem poor. However, it may be that specific arrangements regarding international traffic offer the biggest hope of success. International rail passenger traffic is heavily concentrated on a few key routes mostly characterised by new high speed infrastructure and that can be the subject of negotiations on a case-by-case basis as long as open access is not seen as an important issue. For international rail freight perhaps a regulated simple tariff for that could be achieved even if tariffs for other traffic remain diverse, although the special position of countries such as Switzerland and Austria in international rail freight would clearly have to be taken into account. Whatever the outcome it appears that a pragmatic approach to reducing the current problems in this area will need to be pursued.

Institute for Transport Studies
University of Leeds

Rail Infrastructure Pricing

Key issues and experience from Britain

Chris Nash

1. Introduction

This paper will first consider the particular characteristics of the rail sector which need to be taken into account in any discussion of pricing policy in that sector. It will then comment on key issues in rail infrastructure pricing. Following this a brief review is offered of British experience, particularly relating to the current regulatory review.

2. Characteristics of the rail sector

The rail industry has long been recognised as one whose average costs decline as output increases, ie a declining cost industry. The nature of these declining costs has, for some time, been the subject of detailed study within the literature, leading to the traditional view - that declining costs are a result of economies of scale - being challenged. The distinction between economies of scale, of density and of scope are drawn as follows:

- Economies of scale - as the overall volume of traffic and route network of the railway increase proportionately, the unit costs decline because for instance of spreading of overheads;

- Economies of density - as the volume of traffic between any two points increases, so the unit costs decline, both because typically the result is more intensive use of the infrastructure, and because the operation of longer, more frequent or more direct trains becomes possible;

- Economies of scope - the costs of providing two different types of service, e,g, passenger and freight, by a single operator are less than providing the two by separate operators, because for instance of joint use of infrastructure and rolling stock.

The work of Caves, Christensen, Tretheway and Windle (1985) was an important milestone. Building on work by Harris (1977) and by Keeler (1974), they used data for 43 US railways over the period 1951-1975 to estimate the generalised translog multiproduct cost function. Their estimates found substantial increasing returns to density at every range of output and increasing returns to scale for small carriers, but constant returns for medium and large railroads.

More recent work by Preston (1996) provides further evidence on scale, scope and density economies relating to European conditions. Using data for each of 15 European railways over the period 1971-90 they estimate a series of translog cost functions. Taking into account exclusively operating costs, they find that most companies show increasing returns to density, but only the smaller ones show economies of scale.

Therefore, the phenomenon of economies of density appears to exist for all but the most congested railways. It should not be confused with the issue of whether there are economies of scale when the route network and traffic of a rail company are increased proportionately - the evidence is that beyond some minimum efficient size there are no economies of scale of this nature (Kessides and Willig, 1995).

The result of these inherent railway industry cost structures is that in general short (or indeed long) run marginal cost pricing will leave a deficit which will need to be covered by state subsidy. If the state does not have adequate sources of revenue to cover this deficit, or does not wish to do so for other reasons (e.g. a belief that the resulting redistribution of income will be inequitable, or a fear that such subsidies will encourage inefficiency in production) then some form of second best pricing based on price discrimination and/or multi-part tariffs will be the best option.

In addition, subsidy to rail services is frequently argued for on the grounds that it diverts traffic from other modes which impose greater external costs in the forms of congestion, accidents and environmental pollution. The strength of this additional argument depends on the proportion of any traffic, attracted to rail by the subsidies, that comes from the competing mode and the degree to which charges on that mode are failing to reflect its external costs. It is generally recognised that it would be more efficient to charge the competing mode its full costs than to subsidise rail services for this reason. But if such price increases on competing modes are practically or politically impossible then there is a justifiable argument for rail subsidies on these grounds.

But there are added complications. Firstly just because some subsidies may be justified by efficient pricing does not mean that in practice all subsidies actually given are justified. As suggested above, there is a wide spread view that railways are frequently inefficient both in the production and in the marketing of their services, and that many of the services they provide in less dense areas would be more efficiently provided by another mode. A major reason for these problems is thought to be the way in which railways are controlled and subsidised by the state. The result of this conclusion is a trend towards reform of rail organisation, including separation of operations from infrastructure, increased access for third parties to provide rail services over the shared infrastructure, franchising of some services by means of competitive bidding and in some cases privatisation. Any proposals regarding rail pricing will be unlikely to be acceptable unless they take full account of these trends.

3. Rail infrastructure Pricing

3.1 Overview of Objectives

A number of different objectives for infrastructure charges may be identified. A typical list, drawn from those put forward at the ECMT's 107th Round Table on Transport Economics (ECMT Round Table 107, 1998). which was devoted to the issue of User Charges for Railway Infrastructure, would be:

- promoting efficient use of the infrastructure
- promoting efficient investment in and development of the infrastructure
- recovering the costs of providing the infrastructure, including adequate funding for investment.
- promoting efficiency of operators, for instance through facilitating competition
- harmonisation of the terms of competition between modes

Whilst it may be true that all these objectives may be reduced to single one, that is maximising the social surplus from rail infrastructure (or in other words the difference between benefits and costs valued in money terms), there remain trade-offs between them in terms of the contribution they make to this and not all these objectives can be adequately fulfilled with a single policy instrument. At the ECMT's 107th Round Table there was general agreement that the most important objective was efficient use of the infrastructure, although this should be achieved in the way which least damaged other objectives, in terms for instance of incentives for efficient development of the network and the scope for promoting competition amongst train operators.

In the rest of this section we expand on these objectives drawing on the discussion in ECMT Road Table 107.

3.2 Efficient use and development of infrastructure

The basic principles for the efficient use of infrastructure are that, in the absence of capacity constraints, operators willing to pay the extra costs they impose by their use of the infrastructure should be allowed to use it, whilst in the presence of capacity constraints the capacity should go to the operator and type of traffic for which it has the most value. This of course does presupppose that what the operator is willing to pay represents the social value of the train, so that any external benefits or costs have already been taken account of by taxes or subsidies from the government.

This approach to pricing is essentially that labelled by economists as short run marginal cost pricing; in other words charging the incremental cost of use of the existing infrastructure by the train concerned. This would cover the wear and tear cost, plus any costs imposed on other services in terms of delays or retiming to accommodate the train concerned. In the presence of a capacity constraint, this cost would have added to it the value of any train which could not be run as a result of lack of capacity.

This concept is often contrasted with that of long run marginal cost, which represents the additional cost of an extra train when the infrastructure is optimally adapted to the demand in question. It is well known that if the infrastructure were optimally configured, the two concepts would give the same resulting value, since the infrastructure would be improved to the point at which the cost of the extra capacity exactly matched its value in terms of relieving congestion and permitting additional trains to run. The general perception that short run marginal cost is below long run is only true in the presence of excess capacity; the reverse is true when capacity is scarce.

In practice, indivisibilities and the time lags involved in adapting infrastructure to volume mean that differences between short and long run marginal cost are likely. In this case, some round table participants argued that the theoretically correct approach is to price at short run marginal cost, whilst adapting the infrastructure in accordance with the outcome of social cost-benefit analysis of alternative schemes. In this way, the optimal use of existing infrastructure can be guaranteed, whilst over time, the quantity and quality of infrastructure would be optimised, and the price adapted accordingly. It is also arguably easier to measure short run marginal cost than long run, which - in the presence of indivisibilities - may only be approximated as the average incremental cost of specific capacity enhancing measures which may vary greatly in cost from place to place. However, strong arguments in favour of the long run marginal cost pricing approach were also put forward. Specifically it may give a value which is more stable over time, not fluctuating with day to day changes in the level of congestion, and thus aid planning by the train operating company. Linked to this is the fact that many operators, both freight and passenger, seek contracts running for a number of years in order to justify specific investments in rolling stock or fixed equipment such as terminals. One solution might be to charge long term contracts on the basis of long run marginal cost, but to sell paths on the ‘spot’ market at short run marginal cost.

Another problem of short run marginal cost pricing is that it makes it more profitable for a commercial rail infrastructure company to constrain capacity to force the price up than to invest in expansion. At least if it is regulated to charge long run marginal cost this incentive is removed, although it is still not clear that the appropriate incentive to invest will exist. It may be far more appropriate therefore if long run marginal cost pricing is seen as part of a long run contract which also specifies the infrastructure investment to take place, although such contracts may be difficult to negotiate if several operators are involved.

It may also be doubted whether it really is the case that short run marginal cost is easier to measure than long, particularly in the context of capacity constraints, where alternative operators’ valuations of the slots concerned must be appraised. Alternative ways of doing this were discussed, such as an auctioning system, or permitting individual negotiations over the rate to be charged on the basis of ‘willingness to pay’.

Thus there was a clear consensus that some form of marginal cost pricing was the appropriate basis for rail infrastructure charging but less consensus on the details of its implementation. In point of fact, the Commission appears to have adopted a compromise whereby short run marginal cost is the basis of charging, but the average incremental cost of new capacity rather than the opportunity cost of the slots should be the basis of charging for capacity constrained sections. This approach was recommended in NERA (1998).

3.3 Budget constraints

It was generally agreed - as discussed in section 2 - that railways are subject to economies of traffic density such that any form of marginal cost pricing will typically fail to recover the total cost of providing, maintaining and operating the infrastructure. One solution to this was a contribution from the state. However, some saw this is dangerous in terms of the incentives to efficiency; others feared that it might prove inadequate to fund an appropriate level of investment. In any event it was generally accepted that government funds had a shadow price above one, so that it was appropriate to seek to recover more than simply the marginal cost of infrastructure use from rail operators.

Alternative measures to achieve this with the least possible damage to efficiency were proposed. The standard Ramsey pricing argument would justify raising price above marginal cost in inverse proportion to the elasticity of demand for the service in question. However, it was recognised that it would be difficult to do this in a fixed tariff for more than a limited number of categories of train. Much finer differentiation would be possible if individual negotiations between infrastructure provider and train operator were permitted. It should also be pointed out that the application of Ramsey pricing to an intermediate good is not straightforward; it is the effect on the prices and service patterns in the final market that matter, and that is difficult to predict and appropriately allow for.

The generally advocated alternative to Ramsey pricing is two part tariffs (of course, the two may be combined as well). The attraction of two part tariffs is that the fixed part may be related to ability to pay, but still leave the operator free to raise the necessary cash in the way that loses them the least traffic, without the distorting effect on service levels that a surcharge on the charge per train kilometre has. The difficulty is that if the fixed part is the result of a tariff, it almost inevitably favours large operators against small (even if there is a fixed charge per route kilometre, as in France and Germany, it favours the operators who have a lot of traffic on the particular route, although it is not as damaging to the prospects of entrants as a large fixed charge for an entire network, as in Britain).

The accepted theoretical solution to the problem is the efficient component pricing rule of Baumol (1983). The entrant who comes in on a small scale should be charged marginal social cost plus whatever contribution to the fixed charge the existing operator loses as a result of the new entry. That however is very difficult to estimate.

3.4 Promoting competition

The generally advocated position regarding promoting competition is that tariffs should be simple, transparent and not discriminate between operators. This immediately conflicts with some of the earlier considerations, which suggested that charging for peak capacity, investment and price differentiation may all be best implemented by means of individual negotiations between infrastructure provider and train operator.

It seems that this is the area in which the most obvious trade off is to be made. If it is really seen as crucial to have a simple transparent tariff to promote new entry, then a good deal in terms of efficiency of the use and development of the network may have to be sacrificed. On the other hand, if new competition is less of an issue, perhaps because competition for the market in the form of a franchising system is seen as the most effective way of promoting efficiency in operations, rather than competition in the market, then perhaps simplicity and transparency are not so important.

One suggestion was that if there were any discrimination, it should favour new operators, as they face other barriers to entry in any event. Making paths available at short run marginal cost to new operators might therefore be appropriate. This is consistent with offering paths at short run marginal cost on the spot market, whilst negotiating a higher rate for longer contracts, although in this case the benefit new operators received from a lower price might be outweighed by the increased uncertainty about long term access.

The Commission appears to have concluded that it is in the freight market that new entry is most important and that rail is in the weakest competitive position. Thus it proposed initially that for freight, a strictly marginal cost pricing approach should be adopted with no supplementary measures to raise additional revenue. However, the Council of Ministers has agreed that supplementary charges may be permitted provided that they do not significantly reduce the volume of rail freight.

3.5 Harmonisation across modes

A number of speakers at the Round Table spoke of the need to harmonise pricing systems between modes to achieve a level playing field. That is of course an important efficiency objective, and a failure to follow efficient pricing rules on competing modes will certainly influence the appropriate pricing rule for rail. One confusion should be avoided however. It is not the case that a failure to follow efficient pricing on road or air transport leads to a case for adopting the same inefficiencies on rail, and it is certainly not the case that one should aim for the same relationship between revenue and total cost on all modes. The ratio of marginal social cost to average cost differs between modes and so would the ideal ratio of revenue to total cost. Distortions on one mode certainly lead to a case for varying the price on other modes from marginal social cost, but it will only be optimal for the ratio of the prices to equal the ratio of marginal social cost if there is a fixed amount of traffic to allocate between the modes. This is certainly not true of passenger traffic although it may be a more reasonable approximation for freight. For passenger traffic there is a trade off. Lowering price on rail to offset underpricing on road or air will have the benefit of attracting passengers to divert, but a cost in terms of generating new passengers who value the benefits at less than the marginal social cost. The higher the ratio of generated traffic to diverted the closer the price of rail should stay to marginal social cost.

It will be seen from the above, that whilst greatly illuminating the issues, theoretical considerations alone cannot resolve the question of the most appropriate approach to charging for the use of rail infrastructure. Much depends on the institutional setting in terms of ownership, competition and regulation, and there are still trade-offs to be made which will vary from case to case. This much was generally agreed however. Firstly, some form of marginal social cost pricing should form the basis of the pricing system. If it does not considerable inefficiencies will result. Secondly, this cost should either take the form of long run marginal social cost, including the average incremental cost of creating new train paths, or it should include the congestion and opportunity costs of adding traffic to the existing capacity. The error of associating marginal social cost with simply marginal wear and tear costs should be avoided. Thirdly, it will usually be the case that optimal prices will fall short of covering total cost. Any remaining revenue needs not met by the state, should be satisfied in the least distorting way possible, by means of two part tariffs or price differentiation.

4. British Experience

When the British rail system was privatised, the infrastructure was placed in the hands of a separate company, Railtrack, which was privatised. Passenger operations were franchised out for periods of 5-15 years, whilst freight operations were privatised outright. There is complete open access for freight operators and a limited degree of open access for passenger operators (OPRAF, 1996).

The infrastructure pricing system implemented in Britain for passenger franchises relies on a two part tariff, which involves a variable charge based solely on wear and tear and where appropriate electric traction costs, and a large fixed element based on avoidable costs and an allocation of joint costs (ORR, 1994). Given the arbitrariness of the latter allocation, the system does not necessarily provide good information on the relative profitability of different services. Moreover, the system has been criticised for the very low variable element in the charges. The variable charges include no element either to allow for increased probability of delays caused to other operators or for externalities such as air pollution. From the point of view of efficiency, the system may also reasonably be criticised as having no mechanism to ensure efficient use of scarce capacity. Adjustments in capacity or quality may be made by negotiation between Railtrack and the operators; these however involve major transactions costs. Similarly any requirements for access beyond the 'rights' that accompanied the franchises, as well as all access arrangements for freight or open access passenger operations, are subject to negotiation with Railtrack. There is an independent rail regulator whose role is to approve all access agreements, and he can intervene if he considers that Railtrack is abusing its monopoly power.

A major concern early in the process of privatisation was with the quality of service the monopoly infrastructure operator would provide. In practice, track access agreements include penalty payments for failure to perform, both for Railtrack, where it is to blame, and for the train operating companies where they are responsible. Again there has been much argument as to whether the penalties are adequate.

As part of the Regulatory Review, Railtrack has brought forward evidence for a higher variable element in the charges. This is based on a number of factors:
- engineering evidence that the wear and tear element of the charges does not fully recover these costs
- evidence on the impact on delays to other trains of adding additional services to the system
- an argument that Railtrack needs an incentive payment to encourage increased use of the system, and that such an incentive payment will give it reason to undertake small capacity enhancing investments without costly negotiations over who will pay for them.

The final conclusion of the Regulatory Review will not be published until September, but it appear (ORR, 2000) that the Regulator accepts the broad arguments, whilst differing from Railtrack on many specific points regarding how the new system will be implemented.

Thus Britain will move to a system where the charge for using the infrastructure will consist broadly of a variable element reflecting wear and tear (and electric traction where relevant), a capacity charge reflecting the likely delays imposed on other services, and a fixed charge in the case of the franchisees based essentially on train kilometres. These charges will apply to most modest changes in services, although individual negotiation will still be needed for major projects affecting capacity or quality.

5. Conclusions

We have seen that there is a major conflict between the objectives of rail infrastructure charges and any practical solution involves a trade-off between them. The initial British system for rail passenger franchisees provided a simple tariff in the case of existing rights, whilst leaving all other charges up to individual negotiation. In practice, neither of these approaches has proved adequate. The simple tariff did not adequately reflect the costs of increasing rail services. Negotiations proved costly and complex particularly in the case of relatively small changes which simply might not be worth the effort.

Britain is therefore moving towards a somewhat more complex tariff which comes much closer to marginal cost pricing, in that it allows for a capacity charge related to the expected delays imposed on other trains. However, there will still be large fixed costs for franchisees. Also, there will still be no real 'scarcity fee' related to the opportunity cost of paths when it is simply not possible to run additional services (in this case, apparently the delay cost avoided by removing one existing service will be the basis of the charge. This is likely to understate scarcity values). The role of individual negotiation will be reduced.

The result will be perhaps the most appropriate structure of charges yet seen in Europe. But it still has significant problems. Foremost of these is the treatment of open access operators. Will they be allowed to run paying only the marginal charges and bearing no share of the fixed costs, thus giving them an advantage relative to incumbent operators? Or will they be required to pay a share of the fixed costs, thus distorting prices relative to marginal costs. Perhaps the approach recommended by Coopers & Lybrand (1998) in their report to UIC and now the basis of charges in Germany, under which the entrant can choice to pay on the basis of the same two part tariff as the incumbent, or to pay the same average cost as the incumbent, is the nearest practical approach to the Baumol rule in this situation. But it would clearly be necessary in this case to disaggregate the fixed charge to individual sections of route rather than leaving it as a lump sum for the entire network served by each franchise.

References

Baumol, W J (1983) Some subtle issues in railroad regulation . International Journal of Transport Economics, Vol. 10.

Caves, D W, Christensen, L R, Tretheway, M. and Windle, R J: (1987) Network effects and the measurement of returns to scale and density for US Railroads in A F Daugherty, ed, Analytical Studies in Transport Economics, Cambridge University Press, 97-120.

Coopers & Lybrand (1998) Phase 2 of "Le Peage" study. Final Report. UIC, Paris.

European Conference of Ministers of Transport (1998) Round Table 107 User Charges for Railway Infrastructure. OECD, Paris

Keeler, T A: (1974) Railroad costs, returns to scale and excess capacity. Review of Economics and Statistics, 56, 201-208.

Kessides, I N and Willig, R D (1995) Restructuring regulation of the rail industry for the public interest. World Bank Working Paper 1506, Washington DC.

Harris, R G (1977) Economies of traffic density in the rail freight industry, Bell Journal of Economics, pp556-564.

NERA (1998) An Examination of Rail Infrastructure Charges, NERA, London

Office of Passenger Rail Franchising (1996) Passenger Rail Industry Overview, OPRAF, London

Office of the Rail Regulator (1994) Railtrack's Track Access Charges for Franchised Passenger Services: developing the structure of charges. ORR, London.

Office of the Rail Regulator (2000) The Periodic Review of Railtrack's Access Charges: Provisional Conclusions on the Incentive Framework. ORR, London.

Preston J. (1996) The economics of British Rail privatisation: an assessment. Transport Reviews, 1996, 16 (1) 1-21.

Roy, R. (1998) Infrastructure Cost Recovery under Allocatively Efficient Pricing. UIC, Paris

Infrastructure charges in Europe

A missed chance to increase competitiveness?

G. W. Gustafsson
Infrastructure Director
UIC
Paris

A. Knibbe
Senior Manager
PriceWaterhouseCoopers
The Netherlands

The European railway sector is in danger of being marginalised. The success of the High Speed Network not withstanding, most of the other parts of the railways, and Freight in particular, have not been able to keep up with the growth of the road traffic. Even though the analysis in simple market share terms is misleading, since the bulk of the growth is in the short distance, it is undeniable that the diminishing market share is a threat for rail as a central player.

Through several well known directives the European Union has brought about fundamental change to the rail industry. The separation of infrastructure and operators is amongst the most notable (directive 91/440). This separation however, especially where this is institutionalised, immediately requires a policy on pricing and charging for the use of infrastructure by old and new operators.

By pricing we usually mean the level of the prices imposed on the operators, normally linked with the cost recovery target of the infrastructure manager (IM), such as marginal, marginal social, or full cost recovery. On the contrary by charging we usually mean the way in which the price is calculated: the activity variable, the parameters, discounts, etc.

In the recent “Infrastructure Package” proposal, the EU advocated marginal social cost pricing, at least for freight, and gave some guidelines for charging. On the basis of the most recent proposals of the European Commission, now re-baptised “Railway Package” it is now however clear that little or no pressure to harmonise or standardise pricing and charging across Europe will flow from that. This may not necessarily be expected for charging but should certainly be considered for pricing.

Is this a missed chance to increase the competitiveness of rail? That is what we want to analyse in this article.

We will do this by examining four questions:

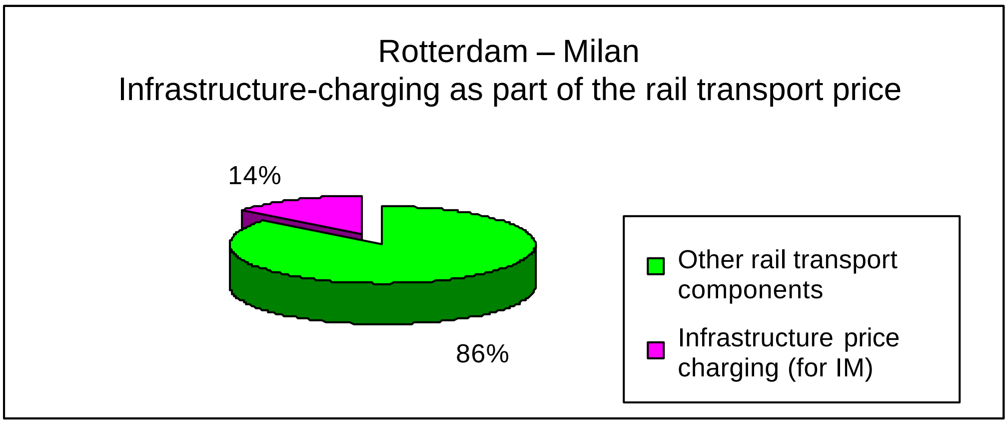
- What is the importance of infra charges?
- What is the height of infra prices in the different European countries?
- Why do charges differ?
- Why should this be a reason to worry about?

Finally we will sum up the conclusions and give some thoughts about the future.

The importance of infra charges

There are many ways to estimate the importance of infra charges. We will deal with this matter in two ways. First we will take a look into the percentage of the price of the RU which he needs to spend on the infrastructure charge. We will do this by using the busy corridor Rotterdam – Milan, and by using a Trans European Rail Freight Freeway (TERFF) slot[1]. In this way we will gain insight to the importance of the charge for freight traffic that only uses the rail mode.

The outcome is that 14% of the selling price of the RU is needed to cover for the infrastructure price, as we can read from the picture below. Other corridors will give other outcomes. In this case the outcome is relatively high, because the prices in Germany are quite high compared with the rest of Europe, since they are based on full cost recovery.



Of course the rail mode is just one of the transport modes used. It is therefore interesting to see what the relative share is of this same infrastructure charge for an intermodal transport with 50 kilometres of pick up and 50 kilometres of delivery at each end. The picture below shows the division of the total intermodal price into pick up and delivery, terminal handling and rail transport costs. Rail transport costs make up 48% of the total intermodal price, and the infrastructure part of the total price is 7%.

It can thus be concluded that the infrastructure price constitutes the smallest part of the total price and that the non-rail costs constitute a substantial part of the total price.

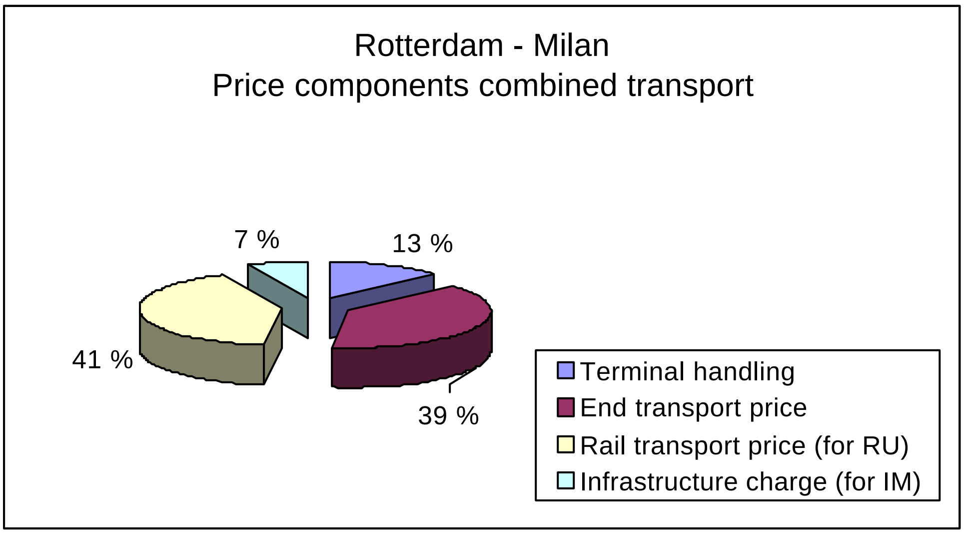
The proportion of the components of the total transport price may vary with local transport and logistic conditions, but calculations on the Le Havre - Sopron and Antwerp - Vienna corridor confirm the thrust of these results.

The RU is a price taker in the market of intermodal transport services on the one hand and on the other hand it can influence the amount of infrastructure price it has to pay to

[1] All the assumptions and all the presented outcomes and more can be found in the report PricewaterhouseCoopers made for the UIC with the title: A comparison of road and rail prices in two international corridors. The final report was submitted in April 2000.

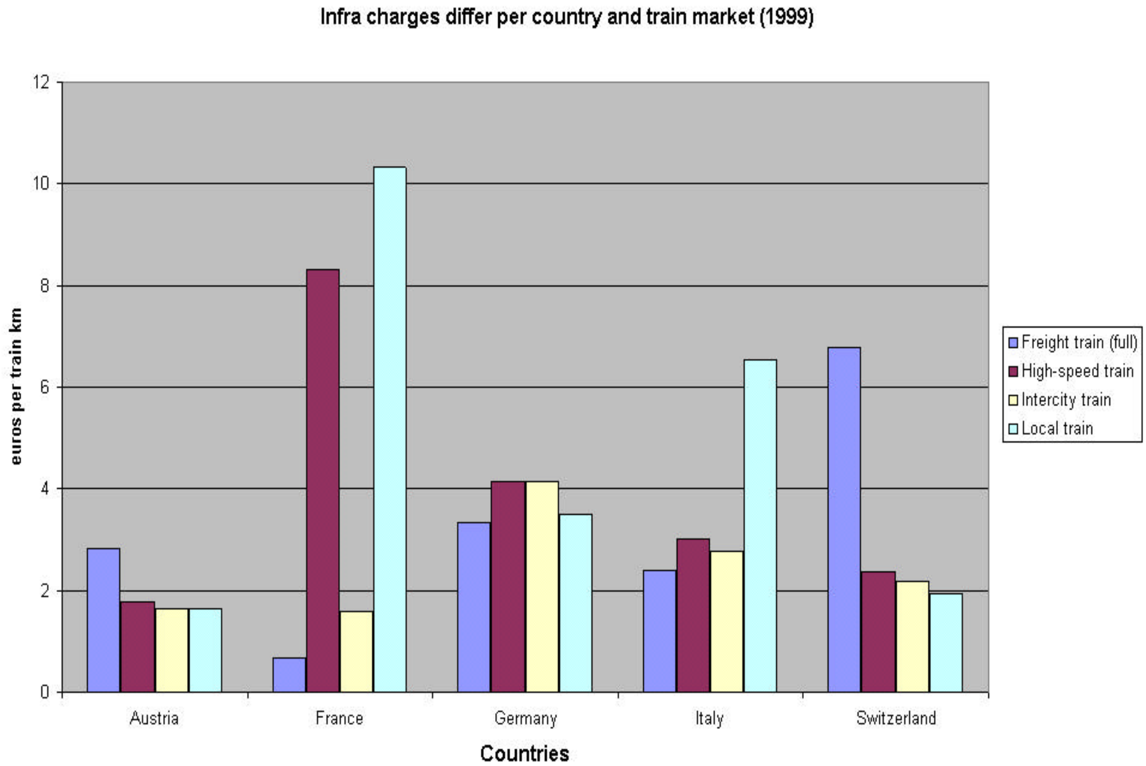
the IM only to a limited extent. Consequently, the RU is stronger on high volume corridors and its position on a low-volume corridor like Le Havre – Vienna is certainly not easy.

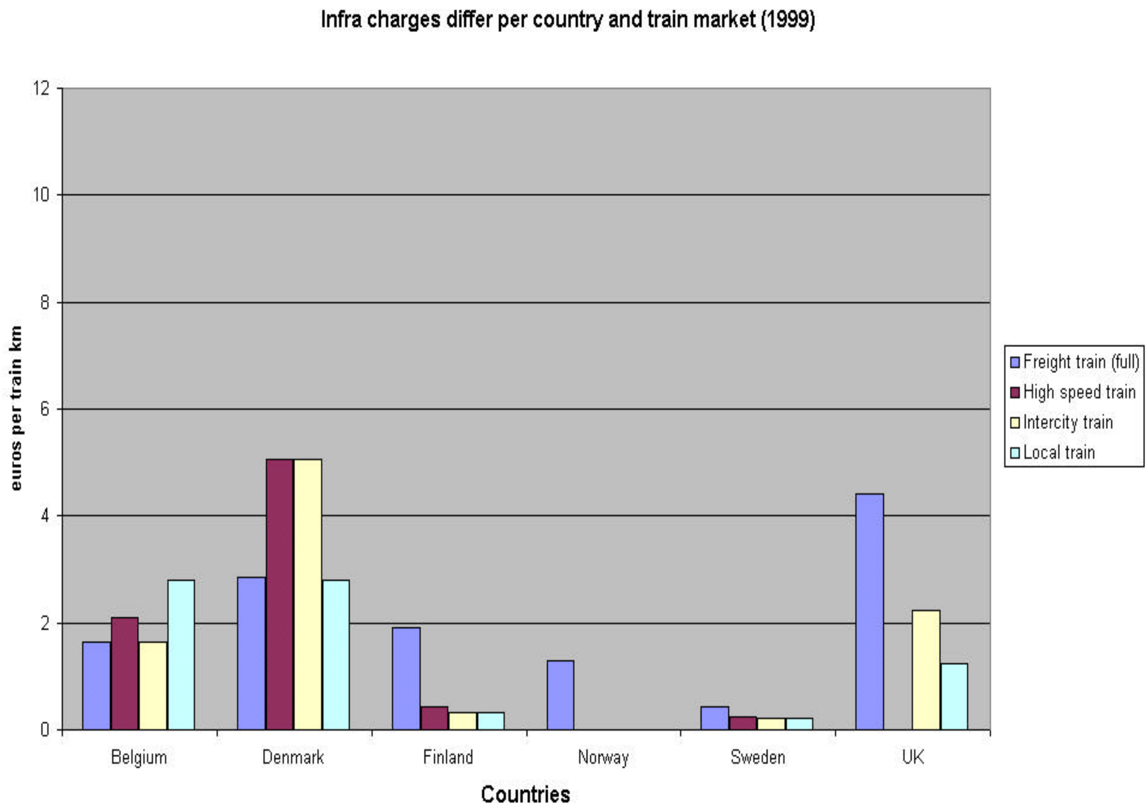
Furthermore the freight RU's in Europe are confronted with high fixed costs and very low margins. The variable costs are very important because a small reduction in those costs will lead to a substantial improvement of the margins. The infra prices are variable costs from the point of view of the RU. In many cases the infra charge will amount to maybe 10 times the margin of a freight RU, so a 10% reduction in the infra charge will lead to a 100% increase in margin. So even if small, the infrastructure charge is nevertheless a highly relevant portion of the price for the RU.



Existing pricing levels in Europe

The existing pricing levels in Europe vary widely. Of course the calculated price depends on many factors. In most countries you will have to specify a specific origin-destination, sometimes you will have to state the time of departure and arrival as well. In Germany and France a two tier charging system is being used. This means that a frequent user of the infrastructure can opt for paying a fixed charge per year for using the network plus a low charge per train kilometre. Infrequent users pay only a much higher variable charge per train kilometre. The result is that frequent users pay a lower average charge per train kilometre than infrequent users. We have made a number of assumptions in order to calculate the real prices to be paid in a number of European countries. In the two following tables the outcomes are presented for 12 countries and 4 types of trains: freight, local, intercity and high speed.


Infra charges differ per country and train market (1999)
euros per train km
0
2
4
6
8
10
12
Austria
France
Germany
Italy
Switzerland
Countries
Freight train (full)
High-speed train
Intercity train
Local train


Infra charges differ per country and train market (1999)
euros per train km
0
2
4
6
8
10
12
Belgium
Denmark
Finland
Norway
Sweden
UK
Countries
Freight train (full)
High speed train
Intercity train
Local train

What is obvious from these figures is that the infra prices vary widely between countries and within countries between types of train. One of the remarkable outcomes is the different treatment of the various train markets. In some countries the freight sector is confronted with higher charges than the passenger services, whereas in other countries the opposite is true. We will not try to explain all the differences but we will give some general reasons for the differences.

Why do charges differ?

There are two main reasons why charges differ between countries:

- different cost recovery levels;
- different charging systems.

In Europe there is a wide range of cost recovery levels, based on different pricing philosophies. For instance in the UK the pricing philosophy is that the IM should charge a fee which allows the IM to make profit. If a subsidy is necessary, the subsidy will go to the railway undertaking that offers this transport service for the lowest subsidy. Railtrack, the IM in the UK, is a stock exchange noted company. Because of the monopoly position of Railtrack, an independent rail regulator is appointed to oversee Railtrack’s actions.

In Scandinavia a completely different philosophy is used. Here the main argument is the level playing field between road and rail. Because road is not paying all the external costs it imposes on the society, like pollution, noise and accidents, rail should be priced lower to make up for this difference. So the charge in these countries is zero (in Norway, which is not a member of the EU) or about the height of marginal social cost. The same is true of the Netherlands.

In several other countries the charge is somewhere in between. The government does give some subsidies to the IM (and to passenger RU’s), but not enough to make it possible for the IM to ask only the marginal costs as infra charge. So the IM’s will have to ask an average charge to make income and expenses balance

The second cause for different pricing levels between the countries is the charging model itself. Even if two countries have the same level of cost recovery they can recover these cost differently. For instance the first country charges freight high and the second country charges passenger high. The first country asks more during peak hours, the second country does not make this difference. The first country has (nearly) all IM services included in the price per train kilometre, while the second country charges all the services separately. From preliminary investigations it seems that all the reasons mentioned are actually valid in today’s Europe[2].

[2] PricewaterhouseCoopers The Netherlands has published a study in November 1999, commissioned by the UIC, in which the details are described of 6 infra charging models in Europe: Infrastructure charging systems for railways.

Why should this be a reason to worry about?

The question arises: why bother about harmonisation of charging models in Europe? At present there is no harmonisation at all. RU's get different incentives in different countries.

Pricing parameters used in Europe
Activity variable
(Ordered) train kilometres
Gross tonn kilometres
Modulating parameters
Average speed of train
Technical features of the track
Axle load of the vehicles
Total gross weight of the train
Traffic density
Peak / off peak
Timetabling margin [1]
Reservation lead time[2]
Relative speed
Required punctuality
Market parameter[3]
External effects (+ and -)
Regularity
Contract period

[1] Defines the slack for drafting the timetable (the lower the required margin the higher the price)
[2] The elapse time between the train path request and either the actual running of the service or the quote / offer by the IM.
[3] A parameter that differentiates between different markets

An indepth study on the infrastructure charging systems of six European countries was conducted late 1999. This research showed that the activity and modulating variables used by the various countries is restricted to the list presented below.

As the table shows for example in some countries train kilometres are being charged and in other countries tonne kilometres. In the first case it is attractive for the RU to use long and heavy trains; in the second it is attractive to frequently shuttle light and short trains. This makes it impossible for the operator to develop a clear cut international production strategy.

To improve the effectiveness of the rail mode it is necessary that RU's receive consistent incentives.

What is even worse if that there is no real transparency of infrastructure charges either. It is nearly impossible to make ''what if'' analyses for RU's: "What if I leave somewhat later and arrive somewhat later, what would that mean for my infrastructure charge in every country I go through?" This is particularly important for freight RU's. Contrary to passenger RU's they have to deal with many last minutes request for transport. Transparency would generate opportunities for RU's to analyse quickly various scenario's in terms of departure and arrival times, travel time and the resulting infrastructure costs. At this moment it takes too much time to obtain and to apply all the necessary information. This is all the more true for potential new entrants of the RU market. Transparency can also lead to shorter response times of IM's, for national traffic, and more importantly for international traffic where more than one IM has to respond quickly. This kind of barriers to entry prevents the rail market from becoming really competitive with other modes. What we actually see is that pricing and charging on a origin-destination basis does not exist; both the price and the charge are an addition of unrelated national based parts. This leads to a patchwork, that in itself is creating a hindrance for more efficient rail transport.

The UIC understood the importance of harmonisation of infrastructure charging systems in an early stage. Two studies were commissioned. Peage I in 1996 made an inventory of the different systems and their elements. Peage II, finished in March 1998, made a proposal how to harmonise the charging models. This harmonisation proposal was put in such a way that it can be implemented without necessarily harmonising the pricing levels for all the countries. A leaflet has been drafted on that basis. Pending the European legislation it has not been adopted yet.

Conclusions

The pricing philosophies for the use of Infrastructure as they have been introduced over the past nine years and the charging models that are applied vary widely and effectively constitute a patchwork where a consistent approach is needed urgently.

Where one would not expect EU legislation to prescribe in detail charging systems the EU, under the pressures of the national governments, has failed to provide the mandatory guidelines to achieve the necessary co-ordination on pricing.

Even though the infrastructure price is a limited constituent of the overall price, it is still an important one given the low profitability of rail freight traffic.

It is clearly in the interest of the European infrastructure managers and the freight operators to take the initiative to implement a truly market and corridor (TERFN!) oriented European charging system. Such a system, consistent and transparent, will certainly increase the competitiveness of rail freight transport.

Helsinki workshop on infrastructure charging on railways

Goals and Principles of Rail Pricing:
Recent Policy Developments at the European Commission

Tom Howes
DG TREN, European Commission
31 July 2000

Introduction: Commission rail policy

European Commission transport policy has its origins in the need to liberalise the European market, historically dominated by public sector monopolies. Liberalisation has progressed at varying speeds in the different sectors of the transport market, but it has reached a critical point in the railway sector.

Commission rail policy has several dimensions, all aimed at revitalising the railways and increasing the efficiency and cost effectiveness of rail services and thence their share of traffic. Currently, there are three dimensions of particular importance. For public transport passenger services, the Commission's intention is to increase the performance and transparency of operations through competition and to this end, just last week, the Commission adopted a regulation on competition for public services for public transport.

The Commission is also striving to create technical and institutional interoperability, with Directives on technical, operational and regulatory interoperability for high speed railways helping to develop the European high speed train network, and with plans for Directives on the interoperability of conventional railways as well.

A third dimension deals with market opening and rail access charging. The current measures contained in the 1998 "railway infrastructure package" address organisational structure (revision of Dir.95/440), rail operator licensing (revision of Dir.95/18), charging, safety, and capacity allocation (revision of Dir.95/19). In terms of the "rail pricing goals, principles and recent developments", it is these last measures which are of greatest interest at this workshop.

Infrastructure charging policy and "the railway package"

This third dimension itself forms the railway element of the Commission's cross-modal transport charging policy of promoting harmonised charging reform based on marginal social costs. This principle was elaborated in the 1998 White Paper on infrastructure charging. It is necessarily a multi-modal approach, as many of the transport problems that have arisen result directly from treating the different modes of transport separately, and of course promotes the gradual harmonisation of charging regimes across Member States as the creation of the single European market remains a key policy goal. The legislation for other modes includes the Eurovignette Directive for roads and the current and planned proposals for airport charging.

The rail package was adopted by the Commission in 1998, and then went to the Council for discussion, and a political agreement was reached in Council in 1999.

The key elements of the railway package agreed in Council include:

- proposals for the further separation of infrastructure manager and operator accounts;
- transparent path allocation requirements and open access to the European Freight network for rail freight operators;
- track access charges based on marginal costs (including environmental costs)
- the possibility of a "mark ups" on the access charge that are not discriminatory and that do not divert traffic which would have paid the marginal costs;
- reservation charges when allocating paths;
- temporary subsidy regimes based on counteracting the effect of unpaid external costs of other modes
- the requirement of independent national rail regulators, with EU co-ordination;

The subsidy regime based on "unpaid external costs" is also contained in the new state aid regulation, adopted last week.

The European Parliament has tabled amendments to the proposed Directives and as the Council of Ministers disagrees with the proposed amendments, the new Treaty of Amsterdam "conciliation" procedure will begin: in mid October Parliament and Council will meet to try to reconcile their differences in the course of the French Presidency.

A regulated European railway charging regime

Assuming that some agreement on the infrastructure package is reached, the European Union will shortly have a regulated rail regime whereby charges should be public, transparent and relate to marginal social costs, where path allocation procedures must be transparent and improve the use of track capacity.

This means that national rail regulators, and any possible Community regulator, who will scrutinise charging regimes, will need to understand and have the means to calculate rail infrastructure, environmental, noise, scarcity and congestion costs and have better path allocation procedures that are transparent and can cope with new entrants.

A lot of the work presented today deals directly with these issues: from the Community side, a considerable body of research has been funded under the 4th and 5th research framework programmes, particularly dealing with cost estimation and path allocation methods. And the work of the High Level Group on infrastructure charging summarises some of this. Future work will include building up the database on costs and path allocation processes. Clearly a number of institutions, not least your universities and research institutes and the national railways, have already begun this task.

Reverting to a multi-modal perspective, there are also a number of related "recent policy developments" including the Commission's forthcoming White Paper on the Common Transport Policy. Dealing with congestion and environmental concerns are the key themes of the paper, as are questions of achieving more modal shift, specifically from road to rail freight and in urban transport patterns.

So for these issues in particular, the development of a state aid regime that allows subsidies for rail freight on environmental grounds, the pressing need for greater investment in rail infrastructure, the ongoing rail charging proposals and the reform of the road charging regime are all clear elements which will be discussed shortly, and form a large part of the solution to Europe's transport problems.

Differential Pricing and Mistake Avoidance

Amihai Glazer
Department of Economics
University of California, Irvine
Irvine, CA 92697

3 July 2000

Abstract

People who suffer disutility from making mistakes may prefer to face consumption choices which reduce their chances of making mistakes. One consequence is that a firm may attract greater demand by charging a single price for different goods than by charging different prices for different goods.

1 Introduction

Economic efficiency calls for charging different prices for goods with different marginal social costs. Yet consumers appear to prefer uniform charges. Train, Ben-Akiva, and Atherton (1989) find that many consumers who must choose among different tariffs for local telephone service often prefer a service with a fixed monthly fee over one with a charge for each call, even if for the number of calls they make the fixed-fee tariff is more expensive than the metered tariff. Internet services, including America On Line, charge a flat monthly fee. This is an interesting case because the firm already is computerized, making its costs of charging by the hour, and varying the charges at peak periods, low indeed. Furthermore, it appears that consumers value some durable goods because they allow consumers to incur a low marginal cost for service. Personal computers and digital cameras have this feature.

Given this experience, it should come as no surprise that governments often ignore economists' advice to vary prices with the marginal social cost of service. For example, a large literature demonstrates that congestion tolls can increase aggregate welfare.[1] Nevertheless, with Singapore a notable exception, congestion tolls are rarely observed.

Experience in the transit industry is similar. In the late 1960s, economists explained how prices that vary with distance and with time of day could increase efficiency. But From 1970 to 1983, only 33 areas introduced time-of-day pricing; 11 of them discontinued the program. For almost all systems studied, over the years inflation reduced the real value of the time-of-day differential; only three areas increased the differential since its inception. And even these numbers exaggerate the use of differential pricing, since in some areas time-of-day fares simplified former fares (Cervero, 1985).

Although it has long been felt that one benefit of electronic fare technology was the greater ease of adopting distance-based pricing, two of the largest transit agencies (Chicago's CTA and New York's MTA) will be implementing flat fares on their electronic fare collection systems. In contrast, 74 percent of bus systems offer passes with unlimited use for a set period (Fleishman, et al., 1996, pp. 24, 28).

More recent data, listed in Table 1, show a continuing decline in use of differential pricing.

[1]For a review see Small (1992).

Table 1: Percentage of Transit Agencies Using Different Forms of Differential Pricing

Year	Peak period surcharges	Transfer surcharges	Zone or distance surcharges
1984	9.5	36.6	34.0
1985	8.6	37.0	33.1
1986	8.8	30.7	27.9
1987	8.4	29.5	33.1
1988	7.8	30.2	33.2
1989	6.4	27.7	31.5
1990	6.5	28.8	38.9
1991	5.5	24.2	39.4
1992	5.6	26.6	39.0
1993	5.6	26.6	39.0
1994	6.4	25.2	37.7
1995	6.5	23.8	36.9
1996	7.0	22.9	32.6
1997	7.0	22.9	32.6
1998	6.1	21.9	32.9

Source: apta.com/stats/fares/fares.htm, from a sample of approximately 300 transit agencies.

Consistent with these data, a survey of 165 transit professionals from 63 US transit systems found 53 percent saying that a very important objective in establishing fares was to "keep fares simple;" only 5 percent thought it "unimportant" (Markowitz, 1986).

This paper explains why consumers may prefer that prices not be differentiated, how a firm can increase profits by charging uniform prices even if its costs increase, and to explore the conditions which make these effects likely to appear.

2 Literature

Some earlier works relate to my topic. Silva (1997) models multiproduct clubs which choose between levying only membership fees, and membership fees combined with separate charges for use of each facility at the club. A different branch of literature explores the strategic effects of spatial price discrimination. Thisse and Vives (1988) consider duopolists competing in a spatially differentiated market. The duopolists play a two-stage game. In stage 1 they choose between uniform pricing and spatial price discrimination; in stage 2 they set prices according to the policy adopted in stage 1. Spatial price discrimination turns out to be a dominant strategy. But profits are lower than with uniform pricing, which puts the firms in a Prisoner's Dilemma.

Choice of price policy has no effect on social surplus in the Thisse-Vives model because individual demand is price inelastic. If demand is price elastic, then equilibria can exist either in which the firms adopt different spatial price policies (De Fraja and Norman 1993) or in which both firms choose uniform prices with the aim of reducing price competition (Aguirre, Espinosa and Macho-Stadler 1998).

The strategic effects of price discrimination have also been studied with nonspatial models of differentiated products. Corts (1998) shows that engaging in third-degree price discrimination can increase or decrease profits, and increase or decrease social surplus. Analogously, Winter (1996) shows that restricting the degree of price discrimination can increase both joint profits and social surplus. Laffont et al. (1998) model duopolistic competition between telecommunications networks. Price discrimination between calls that terminate on a firm's own network and calls that terminate on a rival's network tends to make price competition tougher, while it can increase or decrease surplus. These papers demonstrate that differentiated prices can reduce industry profits and reduce social welfare.

I shall adopt a very different approach, examining the behavior of a monopolist and the psychology of consumers. My approach makes use of regret theory, introduced by Loomes and Sugden (1982). In their model, a consumer chooses an action which has different consequences in different states of the world. A consumer's utility depends, however, not only on the consequence arising from his action, but also on the consequence that would have arisen had he chosen a different action. If the alternative action would have yielded a better consequence, the consumer feels regret.[2]

I assume that a consumer suffers a loss of utility upon realizing that he

[2]See also Loomes and Sugden (1983). Inman, Dyer, and Jia (1997) consider how information about alternatives that were not chosen can affect post-choie evaluation. My paper considers the related problem of pre-choice evaluation. Inman,and Zeelenberg (1998) give experimental evidence that regret affects consumers' choices.

made a mistake, but does not experience added joy upon realizing that he made the correct choice. A similar assumption about asymmetric effects is made in the theory of Loss Aversion.[3]

My paper builds on two central ideas. First, utility may depend not only on the consumption of goods, but also directly declines with the consumer's prior belief that he will make a mistake.[4] Second, and relatedly, people will therefore want to avoid placing themselves in a situation where they may make a mistake, even if this reduces their utility as measured by expected utility from consumption. The reasons for this attitude may be numerous. It may have to do with signaling quality to others (see Scharfstein and Stein (1990), and Boot (1992)). Some consumers may care about revealing information to themselves; for example, they may want to avoid having to memorize facts, for fear that they find they have Alzheimer's disease. Or we may appeal to the evidence that regret theory well explains many phenomena not discussed here. The mistake avoidance drives the result that firms may charge uniform prices, which lead consumers to make fewer mistakes than differentiated prices would.[5]

3 Assumptions

Let a monopolist produce two goods, 1 and 2, with prices p_1 and p_2. For the moment, let the prices be exogenously fixed. Without loss of generality, let $p_1 < p_2$.

There are two types of consumers, A and B. A perfectly informed consumer of type j has a reservation value for good g of r_g^j. To make the problem interesting, suppose that $r_1^A - p_1 > r_2^A - p_2$ (or that a perfectly informed consumer of type A prefers good 1 over good 2), but that the opposite holds for a type-B consumer: $r_1^B - p_1 < r_2^B - p_2$. Also suppose that a consumer will not buy both goods, but that with perfect information he wants to buy one: $r_g^j > p_g$ for all j and for all g. To allow for mistakes, let $r_1^A - p_2 < r_2^A - p_2$:

[3]See Kahneman and Tversky (1979), and Tversky and Kahneman (1992), which refers to the tendency of individuals to weigh losses more heavily than gains.

[4]A simpler model would have a consumer's utility decline whenever he realizes he makes a mistake; that reduced utility could then be assumed to reduce his willingness to purchase the same good in future periods.

[5]See Keasey (1984) for the related idea that consumers may be better off when they are uninformed.

a type-A consumer who mistakenly believes that good 1 has price p_2 will purchase good 2, though under perfect information he would purchase good 1. The fraction of the population that is of type-A is f.

With probability $1 - \mu$ a consumer makes no mistake, correctly observing the price of good 1 and the price of good 2. With probability μ he observes only one price, p_2. The consumer thinks that good 1 may have a price different from p_2, but he does not initially know what that price is. His prior probability on the distribution of p_1 if he makes a mistake is given by the probability distribution function $\phi(p)$. If the consumer seeks to buy good 1 he loses the opportunity of buying good 2, but learns the price of good 1 before he buys it.

The revelation of information is asymmetric. If the consumer does not purchase either good, he does not learn the price of good 1. If, however, he purchases either good, he learns the price of good 1. He learns the price of good 1 either because he has to pay for it, or because consuming good 2 places himself in a situation where prices are made known. (We may think that goods 1 and 2 are train trips to different destinations or at different times. Once on a train, the consumer see there all the prices posted, or perhaps hears from a fellow passenger what are the prices.) The utility loss from making a mistake is Δ.

Of course, with only two goods a consumer who knows that he saw only one price knows he made a mistake. But we can suppose that there a many goods, the consumer not sure exactly how many, but the consumer is interested in only two of them, the ones we consider.

4 Firm's choice between uniform and differentiated prices

I shall first suppose that the firm has a choice between either posting two exogenously set prices, p_1 and p_2, and between posting a single price, $\overline{p}$ for the two goods. We shall see that uniform pricing may lead to higher profits.

4.1 Type-A consumers

Consider first the problem of consumer A. If he observes two, different, prices, then by assumption he made no mistake. Suppose he observed one price, p_2. He know that is the price of good 2, but fears that he may not have noticed the true price of good 1. He may then seek to buy good 1. If the price he observes, p_1, satisfies $r_1^A > p_1$, he buys the good and enjoys consumer surplus $r_1^A - p_1$. But if the price satisfies $r_1^A - p_1 < r_2^A - p_2$, he would have done better by buying good 2. He then suffers a loss of utility of Δ for having made the mistake. His expected utility is then

$$(r_1 - p_1) \int_0^{r_1} \phi(p)dp - \Delta \int_{r_1^A - r_2^A + p_2}^{\infty} \phi(p)dp. \tag{1}$$

Alternatively, he may decide to purchase good 2 at price p_2. He will have made a mistake in purchasing good 2 if $r_1^A - p_1 > r_2^A - p_2$, or if $p_1 < r_1^A - r_2^A + p_2$. Thus, his expected utility if he buys good 2 is

$$r_2 - p_2 - \Delta \int_0^{r_1^A - r_2^A + p_2} \phi(p)dp. \tag{2}$$

The consumer will prefer buying good 2 over good 1 if

$$(r_1 - p_1) \int_0^{r_1} \phi(p)dp - \Delta \int_{r_1^A - r_2^A + p_2}^{\infty} \phi(p)dp > r_2 - p_2 - \Delta \int_0^{r_1^A - r_2^A + p_2} \phi(p)dp. \tag{3}$$

In general we cannot say whether the consumer who thinks he missed the price of good 1 will buy good 2 or instead seek good 1. We do see, however, that

$$\frac{\partial \left((r_1 - p_1) \int_0^{r_1} \phi(p)dp - \Delta \int_{r_1^A - r_2^A + p_2}^{\infty} \phi(p)dp - r_2 + p_2 + \Delta \int_0^{r_1^A - r_2^A + p_2} \phi(p)dp \right)}{\partial p_2}, \tag{4}$$

which is positive, so that for sufficiently high p_2 the consumer would prefer to seek good 1. For sufficiently high Δ, however, expected utility is negative, and so the consumer will not purchase any good at all.

Note that mistake avoidance differs from risk aversion. This can be seen in two ways. First, mistake aversion can lead a consumer to avoid buying good 2, even though he knows for sure its price and even though in the absence of mistake aversion expected consumer surplus is positive and greater than that from good 1. Second, suppose that in the absence of mistake avoidance and in the presence of risk aversion the consumer would prefer to seek good 1 rather than buy good 2. Mistake avoidance can still induce him to avoid seeking good 1, and indeed to avoid entering the market for both goods 1 and 2.

4.1.1 Uniform Price

Contrast these choices to the choice the consumer would make if the firm charges the same price for all the goods, and the consumer knows that it does. That is, instead of a whole list of prices, the firm posts a notice (much like AT&T does for phone calls at 7 cents a minute) that it charges only one price.

Let the firm charge a price $\overline{p}$, with $p_1 < \overline{p} < p_2$. If $r_1^A - \overline{p} > r_2^A - \overline{p}$, and $r_1^A > \overline{p}$, then the uniform price induces the consumer to purchase good 1 (the choice under perfect information), and he purchases the the good even though he would not when prices are differentiated. That is, uniform pricing can both increase the welfare of type-A consumers and increase the profits earned from him—a type-A consumer purchases the good, and pays a price of $\overline{p} > p_1$.

4.2 Type-B consumers

Consider next the behavior of a type-B consumer, who under perfect information would prefer to purchase good 2. The analysis of his decisions is similar to the analysis of consumer A's decisions. If consumer B thinks he may not have noticed the true price of good 1, he fears he can err in two ways: buying good 2 when the price of good 1 would make good 1 more attractive, and seeking to buy good 1 only to learn that he should have bought good 2. Mistake avoidance may then lead consumer B to purchase no good at all.

The assumption that $p_2 > p_1$ means that uniform pricing reduces the price of good 2, which consumer B will buy. Uniform pricing may therefore increase the quantity sold to type-B consumers, but may reduce the firm's revenue from them.

4.3 Profit-maximizing prices

So far the prices under differentiated pricing were considered as exogenously set. Here I extend the analysis by considering the profit-maximizing solution for a monopolist. My aim is to show that a firm may indeed find it profitable to charge uniform prices for different goods.

The number of different cases that can be considered is large. The firm could choose to produce only good 1, only good 2, or both. It could sell only to consumer A, only to consumer B, or to both. To simplify the analysis, and to highlight the optimality of non-differentiated prices, I shall make plausible assumptions that ensure that the firm will produce both goods and sell to both consumers. To that end, suppose that the marginal cost of production is zero for each of the two goods. Let the disutility to a consumer of making a mistake, Δ, be so high that he would not purchase any good if he thought that there is some positive probability of making a mistake. Let $r_1^B = 0$, so that the only good consumer B would purchase is good 2. The maximum price a type-B consumer would be willing to pay for good 2 is r_2^B. For the moment suppose the firm charges r_2^B for good 2. We shall see that can be its profit maximizing strategy.

Recall that for a type-A consumer $r_1^A - p_1 > r_2^A - p_2$, but $r_1^A - p_2 < r_2^A - p_2$. Since Δ is assumed to be large, if consumer A observes only the price of good 2 will not buy any good. Only if he observes both prices, which occurs with probability $1 - \mu$ willl he buy a good. Substituting $p_2 = r_2^B$, we get the maximum price of good 1 that will yet induce consumer A to purchase it as $p_1 = r_1^A - r_2^A + r_2^B$, as long as $r_1^A - p_1 > 0$, that is as long as $r_2^A > r_2^B$. Thus, one possible strategy is for the firm to set $p_2 = r_2^B$ and $p_1 = r_1^A - r_2^A + r_2^B$. We then compare that solution with other strategies. The strategies considered are:

1. The firm charges different prices, p_1 and p_2, as described above. Type-B consumers purchase good 2. Type-A consumers purchase it only if they observe both prices. The firm's revenue is $(1-f)r_2^B + f(1-\mu)\left(r_1^A - r_2^A + r_2^B\right)$.

2. The firm sells only good 1. It charges $p_1 = r_1^A$, and has expected revenue of fr_1^A.

3. The firm sells only good 2. It charges $p_2 = r_2^B$, and has expected revenue of $(1-f)r_2^B$.

4. The firm sells both goods, but charges the same price for each. The maximum price is $\min(r_1^A, r_2^B)$. Since I assumed that $r_2^A > r_2^B$, let $\min(r_1^A, r_2^B) = r_2^B$.

It is easy to see that r_2^B can be greater than fr_1^A and greater than $(1-$

$f)r_2^B$. Consider then the conditions which make

$$r_2^B > (1-f)r_2^B + f(1-\mu)\left(r_1^A - r_2^A + r_2^B\right), \tag{5}$$

or

$$r_2^B > \frac{\left(r_1^A - r_2^A\right)(1-\mu)}{\mu}. \tag{6}$$

This is satisfied for sufficiently high μ, or for sufficiently high probability of making a mistake. That is, the firm's profit-maximizing strategy is to offer two goods at the same price.

5 Discussion

Of course, the explanation given above is not the only one for why consumers may prefer uniform pricing. Risk aversion might explain some cases. As mentioned above, however, if a consumer knows what price he must pay before he purchases a good (as I had assumed), then there is no uncertainty about the price at the time the purchase is made, and therefore risk aversion is irrelevant. But one may think that a consumer knows the price only after he makes a purchase (as often with a telephone calls), or that he must incur a cost, say a fixed search cost, as a stage in consumption, before he knows the prices he will face. But yet the amounts of money involved seem small to evoke much desire for the implicit insurance afforded by uniform prices. Intuition would suggest that the price risks are less important to the rich than to the poor, while casual observation suggests that differentiated pricing is more common in poor countries than in rich ones. And a risk aversion story cannot explain why a visitor to Las Vegas would want to gamble, but would eat lunch at a fixed-price buffet.

My model does not predict that firms will always charge a uniform price. If the disutility of making a mistake is low, then my argument is irrelevant. Or if most consumers prefer one good over the other, then mistake avoidance will lead the firm to provide few types of goods rather than offer many goods at a uniform price. Similarly, if marginal costs of the goods differ, then the firm will not want to offer them all at a uniform price. But the model does highlight additional considerations, and may explain when differentiated prices are most acceptable.

f Fraction of the population that is of type-A

p_g Price of good g

r_g^j Reservation price of good g by consumer j

U_g^j Expected utility of consumer j who purchases good g when he can make a mistake

Δ Utility loss from making a mistake

μ Probability consumer makes a mistake

$\phi(p)$ Prior probability on the distribution of p_1 if consumer makes a mistake

References

[1] Aguirre, Inaki, Maria Paz Espinosa and Ines Macho-Stadler (1998) "Strategic entry deterrence through spatial price discrimination." *Regional Science and Urban Economics*, 28: 297-314.

[2] Boot, Arnold W.A. (1992) "Why Hang on to Losers? Divestitures and Takeovers." *Journal of Finance*, 47: 1401-23.

[3] Cervero, Robert (1985) "Experiences with Time-of-Day Transit Pricing in the United States." *Transportation Research Record*, 1039: 21-30

[4] Corts, K.S. (1998) "Third-degree price discrimination in oligopoly: All-out competition and strategic commitment." *Rand Journal of Economics*, 29(2): 306-323.

[5] De Fraja, Giovanni and George Norman (1993) "Product differentiation, pricing policy and equilibrium." *Journal of Regional Science*, 33: 343-363.

[6] Fleishman, Daniel et al. (1996) "Fare policies, structures and technologies". In Transit Cooperative Research Program, Report 10, Washington, D.C: National Academy Press.

[7] Inman, J. Jeffrey, James S. Dyer, and Jianmin Jia (1997) "A generalized utility model of disappointment and regret effects on post-choice valuation." *Marketing Science*, 16(2): 97-111.

[8] Inman,J. Jeffrey and Marcel Zeelenberg (1998) "'Wow, I could've had a V8!': The role of regret in consumer choice." Tilburg Center for Economic Research Discussion Paper No. 9879.

[9] Kahneman, Daniel, and Amos Tversky (1979) "Prospect theory: An analysis of decision under risk." *Econometrica*, 47: 263-291.

[10] Keasey, Kevin (1984) "Regret theory and information: A note." *Economic Journal*, 94(375): 645-648.

[11] Laffont, Jean Jacqures, Patrick Rey and Jean Tirole (1998) "Network competition II: Price discrimination", *Rand Journal of Economics*, 29(1): 38-56.

[12] Loomes, Graham and Robert Sugden (1982) "Regret theory: An alternative theory of rational choice under uncertainty." *Economic Journal*, 92(368): 805-24.

[13] Loomes, Graham and Robert Sugden (1983) "A rationale for preference reversal." *American Economic Review*, 73(3): 428-32.2

[14] Markowitz, Joel E. (1986) "Prospects for differential transit pricing in the united states." *Transportation Research Record*, 1078: 39-48.

[15] Scharfstein, David S. and Jeremy C. Stein (1990) "Herd behavior and investment." *American Economic Review*, 80(3): 465-479.

[16] Silva, Emilson C.D. (1997) "A-la-Carte or smorgasbord? Multiproduct clubs with costly exclusion." *Journal of Urban Economics*, 41: 264-280.

[17] Small, Kenneth A. (1992) *Urban Transportation Economics*, Vol. 51 of *Fundamentals of Pure and Applied Economics*, Harwood Academic Publishers.

[18] Thisse, Jean Francois and Xavier Vives (1988) "On the strategic choice of spatial price policy." *American Economic Review*, 78(1): 122-137.

[19] Train, Kenneth E, Moshe Ben-Akiva, and Terry Atherton (1989) "Consumption Patterns and self-selecting tariffs." *Review of Economics and Statistics*, 71(1): 62-73.

[20] Tversky, Amos, and Daniel Kahneman (1992) "Advances in prospect theory: cumulative representation of uncertainty." *Journal of Risk and Uncertainty*, 5: 297-323.

[21] Winter, Ralph (1996) "Colluding on relative prices." Working paper, University of Toronto.

Access Pricing for Interconnected Vertically Separated Industries*

A. Bassanini† and Jérôme Pouyet‡

5th October 2000

Abstract

We study the interaction between railroad infrastructure managers in charge of pricing the access to their networks. The infrastructures are used by downstream firms to provide two types of service: domestic and international. The latter requires the use of both networks. Each infrastructure manager must ensure the financial viability of his own network.

We study the Nash equilibrium of the game played by noncooperative infrastructure managers and characterize the strategic interaction between their access pricing decisions. Then, we allow infrastructure managers (or their political principal) to choose to finance the infrastructure's common costs either through a subsidy or solely through user charges. We show that an infrastructure manager sometimes has an incentive to adopt the budget-balance system in order to free-ride on the access prices imposed by the other infrastructure manager.

Keyword: Ramsey Pricing, Interconnected Infrastructures, Financing System.
JEL Classification: L51.

*This paper is part of a joint research program between CERAS-ENPC and IDEI on the economics and the regulation of railroads. Financial and intellectual support from Réseau Férré de France and the Direction des Transports Terrestres (Ministry of Transportation, PREDIT Program) are gratefully acknowledged. We would also like to express our gratitude to Bernard Caillaud and Jean Tirole who made detailed comments and suggestions during this work.

†University of Rome "La Sapienza" and IDEI (Toulouse).

‡Corresponding author. GREMAQ (Toulouse) and CERAS-ENPC, 28 rue des Saint Pères, 75007 Paris, France. E-mail address: pouyet@enpc.fr

1 Introduction

The railway sector in Europe is experiencing a significant reorganization process following the application of EC Directive 91/440. The vertical separation of network management from transport service provision prescribed by the Directive requires a clear definition of the terms of access to railroad infrastructures for downstream transport operators.

The application of the Directive in European Member States has followed different paths, reflecting the heterogeneous nature of the railway networks in the various countries as well as significant differences in pre-existing market structures. In particular, the pricing schemes for infrastructure access vary across countries with respect to the level of infrastructure costs' coverage by users. For instance, the French charging system has enabled RFF to cover about 25% of its total cost, while the percentage is 40% for SCHIG in Austria. On the other hand, the access pricing system implemented by NETZ in Germany has been set with the aim of recovering all costs, excluding those related to new or enhanced infrastructure. Therefore, the role played by infrastructure access pricing can also be markedly dissimilar depending on the objective of the infrastructure manager (or the one of his political principal), or, more generally, on the choice of the mode of regulation.

In view of these differences, and given the Directive's objectives of promoting intra-European traffic, particular attention should be devoted to infrastructure access pricing for international traffic between bordering countries. A proposal for a Council Directive on railway capacity assignment upholds that dissimilar objectives of the infrastructure managers and consequently varying charging systems, lacking acknowledgement of the different markets in which freight services operate and thus inability to grant its international competitiveness require coordination on the side of infrastructure managers in order to avoid heavy impacts on service efficiency and market share.

The aim of this paper is to study the interaction between infrastructure managers in charge of pricing the access to their infrastructures, which are exploited by downstream firms to provide transport services. Our stylized model considers two bordering countries. There are two types of transport services with independent demands, namely domestic and international services. The latter require the use of the railroad infrastructures of both countries. Infrastructure managers maximize national welfare while financing their network deficit.

We argue that a basic distinction should be made according to the type of cost recovery principle adopted. In fact, the fraction of network costs which is not covered by access charges could be funded through taxes levied on the economy as a whole: This kind of approach is named 'taxpayers-pay' financing system. Alternatively, access charges imposed on railway users could be meant to recover the total infrastructure cost: This approach is called 'users-pay' financing system. The main difference is that under the taxpayers-pay financing system the cost of the infrastructure deficit is evaluated at the shadow cost of public funds, whereas under a users-pay financing system it is evaluated at the shadow cost of the budget-balance constraint. Throughout the paper, we will assume that the shadow cost of public funds is smaller than the shadow cost of the budget constraint under a users-pay financing system because of the presence of large fixed costs.

If the infrastructure managers were perfectly cooperating, the optimal access prices would obey standard Ramsey-Boiteux formulas. Moreover, the taxpayers-pay system is socially preferred to the users-pay one since it provides one additional instrument (the subsidy) to the perfectly cooperating infrastructure managers.

However, in an open economy with interconnected infrastructures, the two financing systems entail access charges which, though similar in their Ramsey-Boiteux structure, can differ both in level and strategic nature.

Our results are driven by two basic effects. First, the infrastructure manager of each country will only internalize the fraction of social benefits deriving from international services that accrues to his own consumers. Hence, the international service creates a negative externality on access tariffs: This is the *constituency effect*. Second, the 'perceived' infrastructure cost of the international service for an infrastructure manager differs from the total infrastructure cost of this service because each infrastructure manager is concerned with the financial viability of his network only: This is the *double marginalization effect*. These two effects lead to excessive access prices for the international service.

Let us first mention that with noncooperative infrastructure managers, from the point of view of (total) social welfare, the taxpayers-pay system is even more preferred to the users-pay one. Indeed, under a users-pay system in both countries, since noncooperative access charges are too high, the sum of the shadow costs of the budget constraint in each country is larger than the shadow cost of the budget constraint under perfect cooperation. Therefore, the provision of both the domestic and international services is affected. Consequently, it would be optimal that both infrastructure managers decided simultaneously to finance the infrastructure deficit through a subsidy.

On this basis, we study the strategic choice of the mode of regulation and analyze the individual incentives of the infrastructure managers to adopt one of the two financing systems before setting their access prices. The decision to adopt a particular financing system triggers two effects. First, the change in access prices in a country has a direct effect on the welfare in this country. Second, the change in the international access price in one country entails a modification of the access prices set in the other country; this indirect effect depends on the strategic interaction between infrastructure managers.

A first and obvious result is that, when the infrastructure deficit is sufficiently large, each infrastructure manager prefers the taxpayers-pay system since the access charges needed to balance the budget without subsidy are too high, leading to too large a loss of consumers' surplus. In this case, the direct effect is more important than the indirect strategic effect.

However, an infrastructure manager can sometimes have an incentive to adopt the socially sub-optimal users-pay system. Indeed, consider that, say, country i decides to adopt the users-pay financing system instead of the taxpayers-pay one. For fixed foreign access charges, the taxpayers-pay system is always preferred. Therefore, in this case the direct effect is negative and adopting the users-pay system leads to higher access charges in country i.

Under the strategic substitutability property, when country i increases its access price for the international service, the infrastructure manager in country j is led to decrease his international access charge. This in turn has the following effects: It tends to alleviate the increase in the price of the international service (which benefits to country i's consumers) and to decrease the infrastructure deficit in country i as well as the shadow cost of the budget-balance constraint in this country. The indirect strategic effect is positive when access charges are strategic substitutes. If the infrastructure deficit is initially small, then the positive indirect strategic effect might offset the negative direct effect and the infrastructure manager in country i might prefer to adopt the users-pay financing system because the loss in consumers' surplus might be more than offset by the reduction in the subsidy required to ensure the viability of the infrastructure under a taxpayers-pay system. By contrast, with strategic complements, the indirect strategic effect related to the adoption of a users-pay system is negative. Both effects are then negative and provide each infrastructure manager with the incentive to finance the infrastructure deficit with a subsidy. For two polar cases, we show that depending on (i) the strategic interaction between international access charges, (ii) the level of infrastructure deficit and (iii) the difference between the domestic and the international demands, the incentives to adopt a particular financing system are radically different.

Our paper borrows from two distinct economic literatures. First, we use the extensive works on regulation under a budget constraint, pioneered by Boiteux (1956) and Ramsey (1927) in a different context. This literature has recently been extended to the telecommunication industry, the focus being on the regulation of a vertically integrated industry in which a dominant firm controlling a bottleneck is required to provide interconnection to entrants competing in a complementary segment. For an extensive account of this literature, see Laffont and Tirole (1999). Chang (1996) studies the problem of pricing the access in a vertically separated industry but does not consider the issue of interconnection between infrastructures. Laffont, Rey and Tirole (1998a,b) study the negotiation of access agreements between two networks that need interconnection. Our work differs since we are considering vertically separated industries and our focus concerns more the choice of the mode of regulation. Our model also borrows from the insights obtained by the strategic trade literature, initiated by the seminal paper by Brander and Spencer (1985), in which governments seek to provide their domestic firms with a strategic advantage[1].

The outline of the paper is as follows. In Section 2 we present the model. Then, in Section 3, we study the Ramsey-Boiteux benchmark in which all access pricing decisions are coordinated across countries. This leads, in Section 4, to the determination of the access charges when infrastructure managers behave in a noncooperative way. The nature of the strategic interaction between infrastructure managers is also assessed. In Section 5, we introduce a two-stage game in which, first, infrastructure managers choose the financing system (taxpayers- or users-pay) and, second, determine their access charges; then, we study the equilibria of this game. Finally, Section 6 gathers some concluding remarks. All proofs are relegated to an Appendix.

2 The model

We consider two countries denoted by $i = 1, 2$. In each country an infrastructure manager sets access charges, while downstream firms use the network to provide transport services to final consumers.

The final demand We assume that there are two types of demand, labeled as domestic and international.

Domestic demand corresponds to purely national transport services. Let $q_i(p_i)$ represent the demand function for domestic services in country i (with $\frac{dq_i}{dp_i} \leq 0$) and $S_i(q_i)$ the related net consumers' surplus, with $\frac{dS_i}{dp_i} = -q_i$. All benefits associated to this service accrue to the consumers of country i only.

Similarly, let $q_*(p_*)$ be the international demand for transport services (with $\frac{dq_*}{dp_*} \leq 0$) and $S_*(q_*)$ the related net total consumers' surplus, that is, the net consumers' surplus of *both* countries when a total quantity $q_*(p_*)$ of international services is consumed at price p_*. We then have $\frac{dS_*}{dp_*} = -q_*$ and we assume that country i only internalizes a part θ_i of this surplus. This hypothesis can be justified by assuming that q_* is the total level of round-trip demand for transport (for example, from Paris to Brussels and back to Paris), and θ_i is the fraction of consumers of country i that originates this demand. Then $\theta_1 + \theta_2 = 1$ and the surplus from international services accruing to country i amounts to $\theta_i S_*(q_*)$[2].

[1]Usually, this literature assumes that firms compete in a third market, implying that governments only care about the domestic firm's profit and subsidy.

[2]Other interpretations could easily be thought of. For example, let θ_{ij}^i be the fraction of consumers having a demand for transport from i to j that belongs to country i and q_{ij}^i the related demand. For $i, j = 1, 2$ and $i \neq j$, we have that $\theta_{ij}^i + \theta_{ij}^j = 1$, while $q_{ij} = q_{ij}^i + q_{ij}^j$ is the total demand for international transport from country i to country j. Thus, we are able to define the (net) surplus $S_{ij}(q_{ij})$ related to the demand for international transport. Under the assumption of an isotropic travel pattern $(q_{ij} = q_{ji})$ and with equal prices, we have $S_{ij}(q_{ij}) = S_{ji}(q_{ji})$ and the surplus of consumers in country i related to international transport can be written as $\theta_{ij}^i S_{ij}(q_{ij}) + \theta_{ji}^i S_{ji}(q_{ji}) = \theta_i S_*(q_*)$, where $\theta_i = \theta_{ij}^i + \theta_{ji}^i$ and $q_{ij} = q_{ji} = q_*$.

The infrastructure managers Each infrastructure manager wants to maximize the welfare of his country, which is composed of three terms: (i) the net consumers' surplus, (ii) the infrastructure deficit and (iii) the domestic downstream firms' profits. Let us now describe the last two terms.

To simplify the exposition, we assume that international services travel in each country half of the total number of kilometers[3]. In the absence of subsidies, the profit of the infrastructure in country i is given by

$$\pi_i^{infra} \equiv (a_i - c_u)q_i + (a_{*i} - c_u)q_* - k_i, \tag{1}$$

where a_i and a_{*i} are, respectively, the access charges for a unit of domestic and international transport demand, while c_u is the (constant) marginal cost of the infrastructure in both countries (we assume that it does not depend on the type of service) and k_i is the fixed cost of the network. Notice that for each infrastructure manager the perceived marginal cost for each unit of international demand is c_u, whereas the total marginal cost is actually $2c_u$.

We now discuss an important institutional feature, namely the possibility to use a subsidy to finance the infrastructure deficit. In what follows, we shall consider two possible financing systems:

- Under the 'users-pay' system, the infrastructure manager cannot directly subsidize the infrastructure, and access pricing alone must ensure that infrastructure access charges cover total (fixed and variable) costs. This case is labeled with a superscript 'u'.

- In contrast, under the 'taxpayers-pay' system the infrastructure manager is allowed to finance the infrastructure through taxes levied on the rest of the economy. This case is labeled with a superscript 't'. Taxation is imperfect and has distortionary effects on the rest of the economy. In our partial equilibrium approach, we denote by λ_{pf} the shadow cost of public funds which captures this effect, and we assume that λ_{pf} is the same in both countries.

Denoting by π_i^d the profits of country i's downstream firms, the program of the infrastructure manager in country i will be

$$(\mathcal{P}_{IM_i}^u) \begin{cases} \max\limits_{\{a_i, a_{*i}\}} \left\{ S_i(q_i) + \theta_i S_*(q_*) + \pi_i^{infra} + \pi_i^d \right\} \\ \text{subject to } \pi_i^{infra} \geq 0 \end{cases}$$

under a users-pay system, and

$$(\mathcal{P}_{IM_i}^t) \begin{cases} \max\limits_{\{t_i, a_i, a_{*i}\}} \left\{ S_i(q_i) + \theta_i S_*(q_*) - (1+\lambda_{pf})t_i + \left(t_i + \pi_i^{infra}\right) + \pi_i^d \right\} \\ \text{subject to } t_i + \pi_i^{infra} \geq 0 \end{cases}$$

[3]This entails no loss of generality with respect to the general case where the international traffic travels, say, in country 1 for a fraction α of the total kilometers and in country 2 for a fraction $1-\alpha$.

under a taxpayers-pay system, where t_i is the subsidy to the infrastructure in country i.

The downstream firms' behavior Throughout the paper, we will assume that downstream firms in each sector, namely the two domestic and the international ones, behave competitively[4]. Thus, letting c_d be the constant marginal cost for these firms (we implicitly assume it is the same for both services)[5], the price for the domestic service in country i will be

$$p_i = a_i + c_d. \tag{2}$$

Since we consider round-trip travel, the resulting price in the market for international transport services will be given by

$$p_* = a_{*1} + a_{*2} + c_d. \tag{3}$$

Under the assumption of downstream competitive behavior, downstream firms make no profit and the infrastructure budget constraint coincides with the industry budget constraint (that includes the infrastructure deficit as well as downstream firms' profits).

3 Social optimum and the Ramsey-Boiteux principles for access pricing

As a preliminary to the forthcoming analysis, we consider the benchmark case where the infrastructure managers perfectly cooperate.

We first assume that the unique infrastructure manager adopts a taxpayers-pay financing system. The optimal access charges must therefore solve the following

[4]The assumption of perfectly competitive behavior in the downstream segments significantly simplifies computations.

With downstream market power and linear access prices, there are conflicting forces in presence: on the one hand, an infrastructure manager wants to subsidize at the margin the downstream operators through low access charges to counter their incentive to under-produce and because he cares about the fraction of the downstream operators' profit that accrues to his country; on the other hand, low access prices generate low access revenues. As concerns the international service, the tendency to subsidize the firm might be reduced because (i) each infrastructure manager cares about a fraction of the international surplus and (ii) only a fraction of the downstream operators' profits might accrue to the consumers of his country.

If infrastructure managers can use two-part tariffs, then the fixed part can be used to capture the downstream operators' profits. In this case, another coordination problem arises since, for the international service, both infrastructure managers might be tempted to try to capture the profit made by the downstream operators on this service.

[5]We implicitly assume that the cost of the downstream firms does not depend on travel length. Our setting could be immediately extended to incorporate such considerations.

program

$$\begin{cases} \max_{\{t,a_1,a_2,a_*\}} \left\{ S_1(q_1) + S_2(q_2) + S_*(q_*) - (1+\lambda_{pf})t + \left(t + \pi_1^{infra} + \pi_2^{infra}\right) \right\} \\ \text{subject to } t + \pi_1^{infra} + \pi_2^{infra} \geq 0, \end{cases}$$

where a_* is the unique access charge imposed on the international service. The necessary first-order conditions to be satisfied in interior solutions yield the following optimal access pricing formulas[6]:

$$\frac{p_i^R - c}{p_i^R} = \frac{\lambda_{pf}}{1+\lambda_{pf}} \frac{1}{\eta_i}, \tag{4}$$

for domestic services in country $i = 1, 2$ and

$$\frac{p_*^R - c_*}{p_*^R} = \frac{\lambda_{pf}}{1+\lambda_{pf}} \frac{1}{\eta_*}, \tag{5}$$

where superscript 'R' stands for 'Ramsey'; $c = c_d + c_u$ and $c_* = c_d + 2c_u$ are the social marginal costs of the domestic and international services respectively, and $\eta = -\frac{p}{q}\frac{dq}{dp}$ denotes the elasticity associated to demand $q(p)$. These formulas exemplify the Ramsey-Boiteux recommendations for access pricing.

We would have obtained analogous formulas had we assumed that the (unique) infrastructure manager adopts a users-pay financing system. In this case, denoting by $\tilde{\lambda}$ the shadow cost of the budget constraint, the optimal access charges are given by (4) and (5) in which λ_{pf} is replaced by $\tilde{\lambda}$[7]. Notice that the shadow cost of the infrastructure is now endogenous.

4 The game between infrastructure managers

Let us now consider the situation in which the infrastructure managers act independently. We begin this section with the determination of access charges when both countries adopt the taxpayers-pay or the users-pay financing system, and compare the welfare under the two systems. Then, we determine the nature of the strategic interaction between the infrastructure managers.

4.1 Equilibrium access charges

We will always assume that parameters are such that we obtain interior solutions characterized by first-order conditions.

[6]With interdependent demands similar formulas are obtained except that elasticities are replaced by the so-called super-elasticities. See e.g. Laffont and Tirole (1999).

[7]With cooperation between infrastructure managers, as soon as the total fixed cost is strictly positive, the budget-balance constraint is binding in equilibrium and $\tilde{\lambda}$ is strictly positive.

Taxpayers-pay financing system The program of the infrastructure manager in country i is then given by $(\mathcal{P}^t_{IM_i})$. Standard computations yield the following proposition.

Proposition 1 *The equilibrium access charges under a taxpayers-pay financing system in both countries are characterized by*

$$\frac{p_i^{tt} - c}{p_i^{tt}} = \frac{\lambda_{pf}}{1 + \lambda_{pf}} \frac{1}{\eta_i}, \tag{6}$$

for the domestic service in country $i = 1, 2$ and

$$\frac{a_{*i}^{tt} - c_u}{p_*^{tt}} = \frac{1 + \lambda_{pf} - \theta_i}{1 + \lambda_{pf}} \frac{1}{\eta_*}, \tag{7}$$

for the international service in country $i = 1, 2$.
A sufficient condition for local stability and uniqueness of the equilibrium is

$$(\mathcal{S} - \mathcal{U}) \quad 3\left(\frac{dq_*}{dp_*}\right)^2 - 2q_* \frac{d^2 q_*}{dp_*^2} > 0.$$

The national access tariff is set at its Ramsey-Boiteux level and is therefore optimal. This is intuitive as the infrastructure manager in country i entirely internalizes the surplus generated by this service. However, this also rests on the fact that access pricing in country j does not affect the domestic service in country i because (i) domestic demands are independent from the international one and (ii) the shadow cost of public funds, which gives the social cost of infrastructure financing, is exogenous.

From the perspective of the international demand, matters are different. First, since $\theta_i \in [0, 1]$, each infrastructure manager does not fully internalize the effect of his decision on total international consumers' surplus. The international access charge in country i will thus be excessive: This is the *constituency effect.*

The second effect that guides the pricing of access for the international service in country i is due to the fact that each infrastructure manager does not account for the increase in deficit incurred by the other when he decides to increase his international access charge. With respect to the Ramsey-Boiteux benchmark, the equilibrium price will thus be excessive: This is the *double marginalization effect.*

In order to illustrate our main results, we will later employ specific demand functions. When they are linear or iso-elastic (with an elasticity parameter strictly greater than 1), Condition $(\mathcal{S} - \mathcal{U})$ for local stability and uniqueness is always satisfied.

Users-pay financing system Usually, the taxpayers-pay system and the users-pay system are strongly analogous. However, there is a slight difference that needs to be mentioned.

Under a taxpayers-pay system, the fictitious cost of the budget-balance constraint is the shadow cost of public funds and is given exogenously. On the other hand, under a users-pay system, the cost of the budget constraint is endogenous and depends on the equilibrium configuration. However, we will always have $\tilde{\lambda} \leq \tilde{\lambda}_1 + \tilde{\lambda}_2$, where $\tilde{\lambda}_i$ is the Lagrange multiplier associated to the budget constraint in country i. Indeed, the budget-balance condition as well as the objective of the unique infrastructure manager under cooperation are respectively the sum of the budget constraints and the objectives of the noncooperative infrastructure managers.

Immediate computations show that under a users-pay financing system in both countries the final price is such that

$$\frac{p_*^{uu} - c_*}{p_*^{uu}} = \left(\frac{1 + \tilde{\lambda}_1 - \theta_1}{1 + \tilde{\lambda}_1} + \frac{1 + \tilde{\lambda}_2 - \theta_2}{1 + \tilde{\lambda}_2} \right) \frac{1}{\eta_*},$$

which is therefore larger than the final price under a taxpayers-pay system in both countries. Hence, under a users-pay financing system there is an additional distortion on international services. Moreover, this affects access pricing decisions for domestic services, even though domestic and international demands are independent[8].

4.2 Comparison of financing systems

From the point view of total social welfare, the taxpayers-pay system is preferred to the users-pay system since, beyond the fact that it provides the infrastructure manager with an additional instrument, the non-internalized externalities in the access pricing of international services do not affect domestic charges. Differently stated, given the access charges imposed in country j, the taxpayers-pay system dominates from the point of view of country i, and it does not create a negative externality on country j.

Proposition 2 *From the point of view of total welfare, the simultaneous adoption of the taxpayers-pay financing system is Pareto-superior to the simultaneous adoption of the users-pay financing system*[9].

Let us now introduce another assumption that will be useful for the rest of the analysis.

[8] We cannot compare $\tilde{\lambda}_i$ with $\tilde{\lambda}$. For instance, when $k_i = 0$, we have $\tilde{\lambda}_i = 0 < \tilde{\lambda}$. However, since $\tilde{\lambda} \leq \tilde{\lambda}_1 + \tilde{\lambda}_2$, the sum of the net consumers' surplus associated to the domestic services is larger under cooperation than under noncooperation.

[9] Let us mention that, as noted in Laffont and Tirole (1999) in more general environments, the taxpayers-pay financing system might be 'dangerous' for the following reasons. First, it does not prevent to undertake an undesirable activity (which is financed through taxes on the whole economy). Second, it dilutes the incentives of consumers of the final services to act as watchdogs, that is, to monitor the infrastructure manager's behavior.

Assumption 1 *In each country* $\lambda_{pf} \leq \tilde{\lambda}_i$*, where* $\tilde{\lambda}_i$ *is the Lagrange multiplier associated to the budget constraint of the infrastructure in country* i[10]*.*

Since in the users-pay system the Lagrange multiplier is endogenous, it is quite difficult to come up with a definitive comparison of the two financing systems. This explains the need for this assumption, which seems natural for the railway sector. First, this industry is characterized by the presence of large fixed costs. Second, if this assumption were not satisfied, then we should expect a railway infrastructure manager to finance (part of) the State expenses since raising funds from a distortion in the provision of railway services would have a lower social cost than distortionary taxation on the rest of the economy; this is far from being the case.

4.3 Strategic interaction between infrastructure managers

In this section we analyze the nature of the strategic interaction between the infrastructure managers. In particular, we want to determine how a change in the access prices set in one country (following, say, the adoption of a different mode of regulation) affects the access prices imposed in the other country. In our setting, since domestic demands are independent from the international one, this interaction derives from the access charges set on the international service.

Taxpayers-pay financing system In equilibrium, since transfers are socially costly, the budget constraint will be binding. Replacing the value of the transfer yields the following social welfare in country i

$$SW_i(a_{*i}, a_{*j}) = S_i(q_i) + \theta_i S_*(q_*) - (1 + \lambda_{pf})[k_i - (a_i - c_u)q_i - (a_{*i} - c_u)q_*].$$

The previous equation can be decomposed as the sum of a 'profit' term (the net consumers' surplus, which is affected by the access pricing choice of the other infrastructure manager) and a 'cost' term (the infrastructure deficit, which is also affected by the actions undertaken by the other infrastructure manager). As in a standard IO setting, the sign of $\frac{da_{*j}}{da_{*i}}$ depends on the effect of a marginal variation of a_{*j} on the marginal welfare in country i. Differentiating the welfare function of the infrastructure manager in country i with respect to a_{*i} and a_{*j} we get

$$\frac{\partial^2 SW_i}{\partial a_{*i} \partial a_{*j}} = (1 + \lambda_{pf} - \theta_i)\frac{dq_*}{dp_*} + (1 + \lambda_{pf})(a_{*i} - c_u)\frac{d^2 q_*}{dp_*^2}. \tag{8}$$

[10]This condition implicitly requires that with noncooperative infrastructure managers the budget constraint under a users-pay system in each country is binding in equilibrium. This will be the case whatever the valuation for the international service in country i if the maximal access revenue generated by the international service in country i does not enable to recover the infrastructure fixed cost. In this case, even for a country which does not value the international service the budget constraint will be binding, implying that $\tilde{\lambda}_i > 0$. We make from now on this assumption.

Using the optimality condition on the international access charge in country i (see (7)), we can rearrange (8) to obtain:

$$\frac{\partial^2 SW_i}{\partial a_{*i}\partial a_{*j}} \propto q_* \frac{d^2 q_*}{dp_*^2} - \left(\frac{dq_*}{dp_*}\right)^2 \tag{9}$$

This enables us to state the following proposition.

Proposition 3 *Under a taxpayers-pay financing system, access charges are strategic substitutes (respectively strategic complements) if* $q_* \frac{d^2 q_*}{dp_*^2} - \left(\frac{dq_*}{dp_*}\right)^2 \leq 0$ *(respectively* ≥ 0*).*

The condition on the strategic interaction relates to the log-concavity or log-convexity of the international demand. For instance, if the international demand is log-concave, then access charges are strategic substitutes. Although we do not rule out the case of strategic complements, we will favor the strategic substitutability assumption because (i) log-concavity is more economically appealing and (ii) this condition ensures that the stability-uniqueness condition is always satisfied in equilibrium.

With a concave or linear international demand function ($\frac{d^2 q_*}{dp_*^2} \leq 0$) access charges for the international service will be strategic substitutes whereas for an iso-elastic parameterization ($q_* = p_*^{-\eta_*}$, $\eta_* > 1$) they become strategic complements.

Users-pay financing system In this case, there is an additional difficulty, namely that the Lagrange multiplier associated to the budget-balance condition is endogenous.

However the following simple observation simplifies the analysis: If the access charge for the international service in country j increases, then the profit of the infrastructure in country i decreases, and the budget constraint becomes harder to satisfy. Hence, the Lagrange multiplier associated to the budget-balance condition in country i (which reflects the shadow cost associated to this constraint) is an increasing function of the access charge imposed on the international service in the other country.

Under a users-pay system in country i, the cross-derivative of the social welfare function in country i can be rewritten as follows

$$\frac{\partial^2 SW_i}{\partial a_{*1}\partial a_{*2}} = \frac{1+\tilde{\lambda}_i - \theta_i}{-\frac{dq_*}{dp_*}} \left[q_* \frac{d^2 q_*}{dp_*^2} - \left(\frac{dq_*}{dp_*}\right)^2 \right] + \frac{\partial \tilde{\lambda}_i}{\partial a_{*j}} \left[q_* + (a_{*i} - c_u)\frac{dq_*}{dp_*} \right]. \tag{10}$$

The second bracketed term corresponds to the marginal profit of the infrastructure in country i with respect to the international access charge a_{*j}. The previous condition simply states that under a users-pay system, the strategic interaction must account for the change in the fictitious cost of the budget constraint in country i. Moreover, using the optimality condition (7) we immediately see that in equilibrium

$$q_* + (a_{*i} - c_u)\frac{dq_*}{dp_*} = \frac{\theta_i q_*}{1+\tilde{\lambda}_i} \geq 0.$$

Hence, this implies that with respect to the taxpayers-pay system, access charges tend to be more strategic complements due to the effect on the fictitious cost of the budget constraint. Notice finally that under a users-pay system (weak) concavity of the international demand is no longer sufficient to obtain the strategic substitutability property.

This also has a consequence on the properties of the equilibrium with strategic substitutes. Indeed, in the Appendix we show that under a users-pay financing system in both countries, Condition $(\mathcal{S} - \mathcal{U})$ is also sufficient to guarantee that the equilibrium is unique and locally stable only when access charges are strategic complements. With strategic substitutability, conditions such that the equilibrium is unique or locally stable are much harder to find and are left to future research.

5 Strategic financing of the infrastructure deficit

In this section we study the individual incentives for infrastructure managers to choose one of the two financing systems we consider. We analyze the following two-stage game:

1. The infrastructure managers independently choose a financing system.
2. The infrastructure managers noncooperatively set access charges in their countries.

The outcome of the game will strongly depend on the strategic interaction between the international access charges.

5.1 The direct and the indirect strategic effects

Let us assume that country i has to decide whether to adopt a taxpayers-pay or a users-pay financing system. There are two (sometimes conflicting) effects that guide this decision.

First, consider that foreign access charges are fixed. In this case, the infrastructure manager always prefers a taxpayers-pay system because it provides an additional instrument. This is the direct effect which provides each infrastructure manager with an incentive to adopt the taxpayers-pay instead of the users-pay system.

Notice also that under Assumption 1 the choice of a users-pay system in country i leads to higher access charges in this country (with respect to the taxpayers-pay system). Therefore, for fixed access prices in country j, the adoption of the users-pay system in country i leads to an increase in the infrastructure revenue (since in the relevant domain $\frac{\partial}{\partial a_i}\pi_i^{infra} \geq 0$ and $\frac{\partial}{\partial a_{*i}}\pi_i^{infra} \geq 0$) but to a decrease in the welfare of this country (because consumers' surplus is negatively affected).

The indirect effect depends on the nature of the strategic interaction between international access charges. If access charges are strategic substitutes, then the increase in a_{*i} triggers a decrease in a_{*j}. This variation in the international access price in country j has the following effects in country i: it increases the infrastructure revenue (since $\frac{\partial}{\partial a_{*j}}\pi_i^{infra} \leq 0$), decreases the international price and therefore increases the surplus of consumers in country i, and decreases the shadow cost of the budget constraint in country i. When access charges are strategic substitutes, the indirect strategic effect is positive and makes the adoption of the users-pay system attractive for the infrastructure managers by creating an incentive to save on the subsidy bestowed on the infrastructure and to free-ride on the access prices set in the other country. On the other hand, when access charges are strategic complements the effects are reversed and the indirect effect becomes negative.

5.2 Equilibria of the two-stage game

We now want to determine the equilibria of the two-stage game presented above. In the following, the term 'infrastructure deficit' denotes the amount of subsidy needed to ensure the financial viability of the network under a taxpayers-pay system. Let us start with a first and somewhat obvious result.

Proposition 4 *For a sufficiently large infrastructure deficit, an infrastructure manager prefers to adopt the taxpayers-pay financing system.*

The intuition behind this result is that the shadow cost of the budget constraint becomes large when the infrastructure deficit is large. In this case, the users-pay system entails too large distortions on the access charges[11] and in the unique equilibrium of the two-stage game both infrastructure managers choose a taxpayers-pay financing system, which coincides with the Pareto-optimal outcome.

Since the Lagrange multiplier of the budget constraint is endogenous, it is difficult to determine the general conditions under which a given equilibrium can arise. However, we can obtain a good understanding of the infrastructure managers' incentives by focusing on two polar cases. First, when demands are linear ($q_i(p_i) = \alpha_i - \gamma_i p_i, i = 1, 2$ and $q_*(p_*) = \alpha_* - \gamma_* p_*$), implying that international access charges are strategic substitutes. Second, when demands are iso-elastic

[11] For instance, assume that parameters' values are such that it is impossible to balance the infrastructure deficit without a subsidy. The users-pay system yields a welfare equal to 0 and the taxpayers-pay system is always preferred.

($q_i(p_i) = p_i^{-\eta_i}, i = 1,2, q_*(p_*) = p_*^{-\eta_*}$), so that international access charges are strategic complements. We shall assume that $c_d = 0$[12].

We also assume that the shadow cost of public funds is equal to 0[13]. This implies that (under a taxpayers-pay system) all the access revenue is generated only by the international demand (because domestic access charges are equal to marginal infrastructure costs).

When the infrastructure subsidy is not too large, the equilibria of our two-stage game are markedly different depending on the strategic interaction between infrastructure managers.

Proposition 5 *Assume that demands are linear so that international access charges are strategic substitutes*[14]*. For a small infrastructure deficit in country i:*

- *If country j adopts the taxpayers-pay financing system, then country i prefers to adopt the users-pay system.*
- *Assume that the infrastructure deficit in country j is also small. If country j adopts the users-pay system, then country i prefers to adopt the users-pay system if and only if demands are such that:* $\frac{(\alpha_j - c_u\gamma_j)^2}{\gamma_j} > \frac{(\alpha_* - 2c_u\gamma_*)^2\theta_j}{4\gamma_*}$.

The first part of the proposition exemplifies the incentive of the infrastructure manager in country i to free-ride on his rival: By adopting the users-pay system he increases the infrastructure revenue and triggers a decrease in the international access price in the other country. This reaction is favorable to country i as it further increases the infrastructure revenue and tends to alleviate the increase in the international price and the distortion on the domestic service in this country due to the adoption of the users-pay system. In this case, country i wins from the reduction in its infrastructure deficit whereas country j suffers from the increase in its subsidy. The indirect effect more than offsets the direct effect.

Let us now consider the impact of the valuations for the international service. In the Appendix, we show that the larger the valuation of country i, the larger the incentives of the infrastructure manager to adopt the users-pay financing system. In fact, in our example, the optimality condition (7) for country j can be rewritten as

$$a_{*j} - c_u \underset{\lambda_{pf}=0}{=} (1-\theta_j)\frac{\alpha_* - \gamma_* p_*}{\gamma_*}$$

[12] Given the symmetry of the model, this normalization is without loss of generality.

[13] This assumption is only made for convenience. It could be relaxed but we would obtain much more complex expressions.

[14] As explained in Section 4, the linearity of the international demand in general is not sufficient to have the strategic substitutability property under the users-pay system. In the numerical illustrations proposed in the Appendix one can immediately check that access charges are always strategic substitutes.

Therefore, at the margin, the smaller the valuation of country j, the larger the decrease in the international access price in country j following the adoption of the user-pay system in country i (since the larger was a_{*j} initially). Since the two valuations add up to one, this entails that when country i has a large valuation for the international service, it has a strong incentive to adopt the users-pay system, because the reduction in a_{*j} will be large.

The second part of the proposition shows that when country j adopts the users-pay system there is an additional effect: the increase in the international charge in country i distorts the access pricing decisions in country j through an effect on the shadow cost of the budget constraint in this country. This makes the infrastructure manager in country j less willing to reduce his international charge, especially when he internalizes a large fraction of the international surplus, or when the domestic demand in his country is small with respect to the international one. These two effects combine at the equilibrium[15].

The proof of this proposition (and the following one) relies on the following observation. Assume that country j has chosen the taxpayers-pay system. Then, there exists a value of the fixed cost of infrastructure in country i such that the infrastructure deficit in this country is null. For this value of the fixed cost, denoted by $\underline{k}_i$, the shadow cost of the budget-balance constraint in country i is equal to the shadow cost of public funds and both financing systems yield the same welfare in country i. In a neighborhood of $\underline{k}_i$, we can compute the derivative of the difference of country i's welfare under a taxpayers-pay and a users-pay system. This gives the first part of the previous proposition.

However, when country j adopts the users-pay system, the value of the fixed cost such that both financing systems are equivalent for country i depends on the endogenous shadow costs of the budget constraint in country j (which also depends on country i's first period choice). By assuming that the infrastructure deficit in country j is small, we can get rid of this dependency on endogenous variables (since the shadow costs in country j are almost equal to the shadow cost of public funds). In this case, the infrastructure deficit in country i will be null if the fixed cost of infrastructure in this country is equal to $\underline{k}_i$. Repeating the same procedure, we obtain the second part of the previous proposition which accounts for the effects on the endogenous shadow costs in country j.

If no assumption were made on the infrastructure deficit in country j, then, for the second part of the proposition, we would have obtained a condition which depends on the endogenous shadow costs in country j (and the value of the infrastructure's fixed cost such that both systems yield the same welfare in country i would be dependent on those shadow costs and smaller than $\underline{k}_i$).

Summarizing, in this example it is therefore possible that each infrastructure manager always tries to free-ride on his rival, ending up in an equilibrium in which

[15] It appears to be difficult to clearly separate these effects.

both infrastructure managers choose the users-pay system. In the Appendix, we give three numerical illustrations for each of the possible equilibrium configurations. In particular, multiple (pure strategy) equilibria sometimes emerge[16]: One country chooses a users-pay system whereas the other country sticks to the taxpayers-pay system because it would be too costly to adopt the users-pay system, and conversely.

Figure 1 illustrates the previous proposition in the case $\frac{(\alpha_j - c_u \gamma_j)^2}{\gamma_j} > \frac{(\alpha_* - 2c_u \gamma_*)^2 \theta_j}{4\gamma_*}$; if this condition is violated, then the first zone in which adopting the users-pay system is a dominant strategy for country i disappears.

Figure 1 here

With iso-elastic demands and the strategic complementarity property, the infrastructure managers' incentives are radically different.

Proposition 6 *Assume that demands are iso-elastic so that international access charges are strategic complements. For small infrastructure deficits each infrastructure manager prefers to adopt the taxpayers-pay system.*

In this case, the strategic reaction of country j after country i adopts the users-pay system affects negatively country i's welfare. With respect to the linear demands case, the logic of the argument is reversed since both the direct and the indirect effects are negative. Under the strategic complementarity assumption, the unique equilibrium of the two-stage game involves each infrastructure manager choosing a taxpayers-pay financing system.

5.3 The creation of international corridors

The European Union strongly encourages the creation of corridors as a means to develop international (especially freight) traffic (see EC White Paper "A Strategy for Revitalising the Community's Railways"). Let us note that our previous modeling captures some of the elements inherent to corridor creation. Indeed, each infrastructure manager will have to determine the access charge to be paid for the part of the corridor that concerns his infrastructure. Then, the sum of these access charges will determine the unique access charge for the use of the corridors. Therefore, if infrastructure managers do not succeed in reaching a high level of cooperation, we should expect the international traffic to be excessively charged, due to the constituency and the double marginalization effects and also possibly to socially sub-optimal choices of the mode of regulation.

Indeed, the creation of the corridor between Germany, Netherlands, Switzerland, Austria and Italy has not succeeded in developing freight traffic. According to Le Monde (18/07/2000) the two main reasons for this failure are that (i) access charges

[16] In this case, a mixed strategy equilibrium also exists.

paid in Germany are prohibitively high and (ii) Germany has refused to reserve tracks for this corridor.

The creation of the Belifret corridor (Belgium, Luxembourg, France, Italy and recently Spain) is somewhat more successful. However, another coordination problem arises due to the 'pass-through' nature of the services using this corridor: Some countries that are part of the corridor do not value the surplus associated to international services running on the corridor and therefore tend to lobby in favor of high access charges. This raises another issue, namely that of supra-national cooperation between national infrastructure managers.

6 Conclusions

In this paper, we have studied the interaction between railroad infrastructure managers when an international service requires the use of both infrastructures. We have isolated two effects: The constituency effect, related to the fact that each infrastructure manager only internalizes a fraction of the surplus generated by the international service, and the double marginalization effect, due to the difference between the infrastructure cost perceived by each infrastructure manager and the total infrastructure cost of the international service.

The interaction created by the international service can provide an infrastructure manager with an incentive to adopt a sub-optimal financing system in order to free-ride on the other country and decrease the amount of subsidy bestowed on the infrastructure.

We have remained silent on a number of questions.

For instance, our model implicitly assumes that networks are interconnected: Firms can always go from one country to the other. However, it has been argued that the development of the international traffic also suffers from a poor quality of interconnection. This is the so-called interoperability problem, which appears to be critical for the development of intra-European networks.

Another question concerns the investment decisions undertaken by the infrastructure managers. More specifically, the decisions to create or to close lines should be incorporated in our framework[17], and should not be neutral with respect to the strategic interaction between infrastructure managers or the possibility to choose a particular financing system.

Finally, our model has highlighted many coordination failures between national infrastructure managers. Future work should study the design of supra-national institutions or rules that enable to implement a certain level of cooperation. We leave these extensions to future research.

[17] For instance, the fixed cost of maintaining the line is much lower for an only-freight line because of lower safety standards.

7 References

Boiteux, M., 1956, "Sur la Gestion des Monopoles Publics Astreints à l'Equilibre Budgétaire", *Econometrica*, 24: 22-40. Published in English as "On the Management of Public Monopolies Subject to Budgetary Constraints", *Journal of Economic Theory*, 3: 219-40.

Brander, J.A. and B.J. Spencer, 1985, "Export Subsidies and International Market Share Rivalry", *Journal of International Economics*, 18, 83-100.

Chang, M.C., 1996, "Ramsey Pricing in a Hierarchical Structure with an Application to Network-Access Pricing", *Journal of Economics*, 64: 281-314.

European Commission, Council Proposal COM(98)480 def., 98/0266 (SYN).

European Commission, Green Paper "Towards Fair and Efficient Pricing in Transport", COM(95)691.

European Commission, White Paper "A Strategy for Revitalizing the Community's Railways", COM(96)421 def.

Laffont, J.J. and J. Tirole, *Competition in Telecommunications*, March 1999.

Laffont, J.J., P. Rey and J. Tirole, 1998a, "Network Competition: I. Overview and Nondiscriminatory Pricing", *Rand Journal of Economics*, 29: 1-37.

Laffont, J.J., P. Rey and J. Tirole, 1998b, "Network Competition: II. Price Discrimination", *Rand Journal of Economics*, 29: 38-56.

Le Monde, 18/07/2000, electronic edition.

NERA (1998), "An Examination of Rail Infrastructure Charges", Final Report for the European Commission, London.

Ramsey, F., 1927, "A Contribution to the Theory of Taxation", *Economic Journal*, 47.

8 Appendix

Throughout this Appendix we omit arguments of functions for simplicity.

8.1 Social optimum and the Ramsey-Boiteux principles for access pricing

In equilibrium the budget-balance condition will be binding. This determines the value of the transfer. Optimizing with respect to the access charges and rearranging terms yields (4) and (5).

Using these first-order conditions the second-order conditions of the maximization problem will be satisfied if $(1+2\lambda_{pf})(\frac{dq_i}{dp_i})^2 \geq \lambda_{pf} q_i \frac{d^2 q_i}{dp_i^2}$, $i = 1,2$ and $(1+2\lambda_{pf})(\frac{dq_*}{dp_*})^2 \geq \lambda_{pf} q_* \frac{d^2 q_*}{dp_*^2}$.

In the case of an iso-elastic international demand $q_* = p_*^{-\eta_*}$ the second-order condition amounts to $\eta_* \geq \frac{\lambda_{pf}}{1+\lambda_{pf}}$.

8.2 Equilibrium access charges

Taxpayers-pay in both countries In equilibrium the budget-balance condition will be binding in both countries. This defines the value of the transfer in each country. Then, optimizing with respect to the access charges and rearranging terms yields (6) and (7).

The second-order conditions will be satisfied if

$$(SOC_i) \quad (1+2\lambda_{pf})\left(\frac{dq_i}{dp_i}\right)^2 \geq \lambda_{pf} q_i \frac{d^2 q_i}{dp_i^2}$$

$$(SOC_{*i}) \quad (2+2\lambda_{pf}-\theta_i)\left(\frac{dq_*}{dp_*}\right)^2 \geq (1+\lambda_{pf}-\theta_i) q_* \frac{d^2 q_*}{dp_*^2}.$$

We will always assume that these two conditions are satisfied whatever $\theta_i \in [0,1]$. This implies that an equilibrium exists.

The equilibrium will be locally stable if

$$\frac{\partial^2 SW_i}{\partial a_{*i} \partial a_{*j}} \frac{\partial^2 SW_j}{\partial a_{*i} \partial a_{*j}} < \frac{\partial^2 SW_i}{\partial a_{*i}^2} \frac{\partial^2 SW_j}{\partial a_{*j}^2}.$$

Using the first-order condition with respect to a_{*i} we immediately obtain

$$\frac{\partial^2 SW_i}{\partial a_{*i} \partial a_{*j}} = \frac{1}{\frac{dq_*}{dp_*}}(1+\lambda_{pf}-\theta_i)\left[\left(\frac{dq_*}{dp_*}\right)^2 - q_* \frac{d^2 q_*}{dp_*^2}\right],$$

$$\frac{\partial^2 SW_i}{\partial a_{*i}^2} = \frac{1}{\frac{dq_*}{dp_*}}\left[(2+2\lambda_{pf}-\theta_i)\left(\frac{dq_*}{dp_*}\right)^2 - (1+\lambda_{pf}-\theta_i) q_* \frac{d^2 q_*}{dp_*^2}\right],$$

for $i \neq j$. Direct computations show that the equilibrium will be locally stable whatever $\theta_i \in [0,1]$ if

$$(2+3\lambda_{pf})\left(\frac{dq_*}{dp_*}\right)^2 - (1+2\lambda_{pf})q_*\frac{d^2q_*}{dp_*^2} > 0. \tag{11}$$

A sufficient condition for the equilibrium to be unique is

$$\left|\frac{\partial^2 SW_i}{\partial a_{*i}^2}\right| > \left|\frac{\partial^2 SW_i}{\partial a_{*i}\partial a_{*j}}\right|. \tag{12}$$

Using the first-order condition with respect to a_{*i}, this condition can be rewritten as follows

$$\left|\frac{2+2\lambda_{pf}-\theta_i}{1+\lambda_{pf}-\theta_i}\left(\frac{dq_*}{dp_*}\right)^2 - q_*\frac{d^2q_*}{dp_*^2}\right| > \left|\left(\frac{dq_*}{dp_*}\right)^2 - q_*\frac{d^2q_*}{dp_*^2}\right|.$$

This condition is trivially satisfied when international access charges are strategic substitutes. Under the strategic complementarity property, using the second-order condition with respect to a_{*i} (SOC_{*i}), the condition for uniqueness can be rewritten as follows

$$(3+3\lambda_{pf}-2\theta_i)\left(\frac{dq_*}{dp_*}\right)^2 > 2(1+\lambda_{pf}-\theta_i)q_*\frac{d^2q_*}{dp_*^2},$$

which will be satisfied whatever $\theta_i \in [0,1]$ if

$$3\left(\frac{dq_*}{dp_*}\right)^2 - 2q_*\frac{d^2q_*}{dp_*^2} > 0.$$

Notice that the sufficient condition for uniqueness implies the local stability property.

In the case of an iso-elastic international demand $q_* = p_*^{-\eta_*}$, (SOC_i) amounts to $\eta_i \geq \frac{\lambda_{pf}}{1+\lambda_{pf}}$, ($SOC_{*i}$) amounts to $\eta_* \geq \frac{1+\lambda_{pf}-\theta_i}{1+\lambda_{pf}}$, local stability amounts to $\eta_* > \frac{1+2\lambda_{pf}}{1+\lambda_{pf}}$, uniqueness amounts to $\eta_* > 2$, existence (i.e. positivity of a_{*i}) amounts to $\eta_* \geq \frac{1+2\lambda_{pf}}{1+\lambda_{pf}}$, the infrastructure profit in country i will be a concave function of a_{*i} if $\eta_* \geq 1$.

Users-pay in both countries In this case, we have to account for the fact that the Lagrange multiplier associated to the budget constraint in country i depends on the access charge for the international service set in country j. Using the optimality condition for the international access charge, simple manipulations show that

$$\frac{\partial^2 SW_i}{\partial a_{*i}^2} = \frac{1}{\frac{dq_*}{dp_*}}\left[(2+2\tilde{\lambda}_i-\theta_i)\left(\frac{dq_*}{dp_*}\right)^2 - (1+\tilde{\lambda}_i-\theta_i)q_*\frac{d^2q_*}{dp_*^2}\right], \tag{13}$$

$$\frac{\partial^2 SW_i}{\partial a_{*i}\partial a_{*j}} = \frac{1+\tilde{\lambda}_i-\theta_i}{\frac{dq_*}{dp_*}}\left[\left(\frac{dq_*}{dp_*}\right)^2 - q_*\frac{d^2q_*}{dp_*^2}\right] + \frac{\partial\tilde{\lambda}_i}{\partial a_{*j}}\frac{\theta_i q_*}{1+\tilde{\lambda}_i}. \tag{14}$$

We have a partial result on the Lagrange multiplier.

Lemma 1 *The Lagrange multiplier* $\tilde{\lambda}_i$ *associated to the budget-constraint in country* i *under a users-pay financing system is such that* $\frac{\partial \tilde{\lambda}_i}{\partial a_{*j}} \geq 0$.

This is intuitive. *Ceteris paribus* an increase in a_{*j} decreases the infrastructure revenue and therefore hardens the budget-balance constraint.

We look for a condition such that the sufficient condition (12) that ensure the uniqueness, and consequently the local stability, of the equilibrium is satisfied. Under the strategic complementarity property, we always have

$$\frac{\partial^2 SW_i}{\partial a_{*i} \partial a_{*j}} \geq 0.$$

Using the second-order condition (SOC_{*i}), this implies that (12) can be rewritten as follows

$$\frac{1}{\frac{dq_*}{dp_*}} \left[(3 + 3\tilde{\lambda}_i - 2\theta_i) \left(\frac{dq_*}{dp_*} \right)^2 - 2(1 + \tilde{\lambda}_i - \theta_i) q_* \frac{d^2 q_*}{dp_*^2} \right] + \frac{\theta_i q_*}{1 + \tilde{\lambda}_i} \frac{\partial \tilde{\lambda}_i}{\partial a_{*j}} < 0.$$

From Lemma 1, a sufficient condition to ensure that the last inequality is satisfied is

$$(3 + 3\tilde{\lambda}_i - 2\theta_i) \left(\frac{dq_*}{dp_*} \right)^2 - 2(1 + \tilde{\lambda}_i - \theta_i) q_* \frac{d^2 q_*}{dp_*^2} > 0. \tag{15}$$

Condition (15) will be satisfied for all $\theta_i \in [0, 1]$ if

$$3 \left(\frac{dq_*}{dp_*} \right)^2 - 2q_* \frac{d^2 q_*}{dp_*^2} > 0,$$

which is the same condition we found when both countries adopt the taxpayers-pay financing system.

Under the strategic substitutability property, we cannot conclude because we were not able to sign $\frac{\partial^2 SW_i}{\partial a_{*i} \partial a_{*j}}$.

8.3 Strategic financing of the infrastructure deficit

We use the same methodology for the cases of linear and iso-elastic demands. We study the incentives for the infrastructure manager in country 1 to adopt the users-pay financing system instead of the taxpayers-pay one, given the financing system adopted in country 2. For each case (taxpayers-pay or users-pay in country 2), we start by identifying the conditions on the fixed cost such that the Lagrange multiplier associated to the budget-balance constraint in country 1 is equal to the shadow cost of public funds. Then we compute the difference in welfare in country 1 under a taxpayers-pay and a users-pay system. Finally, we study the derivative of this difference with respect to the fixed cost in country 1 when the Lagrange multipliers are close to the shadow cost of public funds. We will give a numerical illustration of every type of equilibrium we will exhibit. The discussion is couched in terms of fixed costs but only the infrastructure deficit is relevant.

Let us introduce the following notation: $\tilde{\lambda}^i_{ut}$ is the Lagrange multiplier associated to the budget-balance constraint in country i when country 1 adopts the users-pay system and country 2 adopts the taxpayers-pay one. A similar notation is used for country j. From the perspective of the first stage of the game, there are four possible states: (u,u), (u,t), (t,u) and (t,t). We assume that $\lambda_{pf} = 0$ and $c_d = 0$. Note that $\theta_1 + \theta_2 = 1$, so that the country 2's valuation for the international service is $1 - \theta_1$.

8.3.1 Strategic substitutability and linear demands

Let us assume that the domestic demand functions are given by $q_i(p_i) = \alpha_i - \gamma_i p_i$, $i = 1, 2$. The international demand is given by $q_*(p_*) = \alpha_* - \gamma_* p_*$.

Taxpayers-pay in country 2 If country 1 also adopts the taxpayers-pay system, then the optimal access charges are given by (6) and (7). Given that $\lambda_{pf} = 0$ the access charge for the domestic service in country 1 is equal to the marginal cost of the infrastructure c_u and the profit of the infrastructure in country 1 can be rewritten as

$$\frac{(\alpha_* - 2c_u\gamma_*)^2(1-\theta_1)}{4\gamma_*} - k_1.$$

Hence, if the fixed cost of the infrastructure is equal to $\underline{k}_1 \equiv \frac{(\alpha_* - 2c_u\gamma_*)^2(1-\theta_1)}{4\gamma_*}$, then the optimal access charges are such that the infrastructure breaks even without any subsidy. Accordingly, we have $\tilde{\lambda}^1_{ut} = \lambda_{pf} = 0$ and the infrastructure manager in country 1 is indifferent between the two financing systems.

Let us denote by

$$\Delta SW^1_t(\tilde{\lambda}^1_{ut}, k_1) \equiv SW^1_{tt}(k_1) - SW^1_{ut}(\tilde{\lambda}^1_{ut}, k_1)$$

the difference in welfare in country 1 under a taxpayers-pay system and a users-pay system when country 2 adopts the taxpayers-pay system. We have, after simple computations,

$$\frac{d\Delta SW^1_t}{dk_1} = \frac{\partial \Delta SW^1_t}{\partial k_1} + \frac{\partial \Delta SW^1_t}{\partial \tilde{\lambda}^1_{ut}} \frac{\partial \tilde{\lambda}^1_{ut}}{\partial k_1} \tag{16}$$

$$= -1 + (1 + \tilde{\lambda}^1_{ut}) \left[\frac{\omega_1}{(1 + 2\tilde{\lambda}^1_{ut})^3} + \frac{\omega_* \theta_1^2}{\left[2 + (2 + \theta_1)\tilde{\lambda}^1_{ut}\right]^3} \right] \frac{\partial \tilde{\lambda}^1_{ut}}{\partial k_1}, \tag{17}$$

where $\omega_i = \frac{(\alpha_i - c_u\gamma_i)^2}{\gamma_i}$ and $\omega_* = \frac{(\alpha_* - 2c_u\gamma_*)^2}{\gamma_*}$.

When country 1 adopts the users-pay financing system, replacing the optimal access charges as function of the Lagrange multiplier, the budget-balance condition can be rewritten in equilibrium as follows

$$\omega_1 \frac{\tilde{\lambda}^1_{ut}(1+\tilde{\lambda}^1_{ut})}{(1+2\tilde{\lambda}^1_{ut})^2} + \omega_* \frac{(1+\tilde{\lambda}^1_{ut})(1+\tilde{\lambda}^1_{ut}-\theta_1)}{\left[2+(2+\theta_1)\tilde{\lambda}^1_{ut}\right]^2} = k_1. \qquad (18)$$

Totally differentiating (18) and taking the limit when k_1 goes to $\underline{k}_1$ (implying that $\tilde{\lambda}^1_{ut}$ goes to $\lambda_{pf} = 0$), we obtain

$$\frac{\partial \tilde{\lambda}^1_{ut}}{\partial k_1}|_{k_1=\underline{k}_1} = \frac{1}{\omega_1 + \frac{\omega_* \theta_1^2}{4}}. \qquad (19)$$

Finally, replacing (19) in (17) we obtain

$$\frac{d\Delta SW^1_t}{dk_1}|_{k_1=\underline{k}_1} \propto -\frac{\omega_* \theta_1^2}{8} < 0.$$

This enables us to state the following lemma.

Lemma 2 *Assume that demands are linear, $c_d = 0$, $\lambda_{pf} = 0$ and that country 2 adopts the taxpayers-pay system. In a neighborhood of $\underline{k}_1$ country 1 prefers to adopt the users-pay system than the taxpayers-pay one.*

Finally, notice that to obtain this result we do not need to make any assumptions on the infrastructure deficit in country 2.

Users-pay in country 2 There is now an additional difficulty. When both countries adopt the users-pay financing system, a variation in the fixed cost of the infrastructure in country 1 affects the Lagrange multipliers in both countries.

As previously, denote by

$$\Delta SW^1_u(\tilde{\lambda}^1_{uu}, \tilde{\lambda}^2_{uu}, \tilde{\lambda}^2_{tu}, k_1) \equiv SW^1_{tu}(\tilde{\lambda}^2_{tu}, k_1) - SW^1_{uu}(\tilde{\lambda}^1_{uu}, \tilde{\lambda}^2_{uu}, k_1)$$

the difference in welfare in country 1 under a taxpayers-pay and a users-pay financing system when country 2 adopts the users-pay system. We are interested in the sign of

$$\frac{d\Delta SW^1_u}{dk_1} = \frac{\partial \Delta SW^1_u}{\partial k_1} + \left[\frac{\partial \Delta SW^1_u}{\partial \tilde{\lambda}^1_{uu}} + \frac{\partial \Delta SW^1_u}{\partial \tilde{\lambda}^2_{uu}} \frac{\partial \tilde{\lambda}^2_{uu}}{\partial \tilde{\lambda}^1_{uu}}\right] \frac{\partial \tilde{\lambda}^1_{uu}}{\partial k_1} \qquad (20)$$

If country 1 adopts the taxpayers-pay system then there will be no need to finance the infrastructure if

$$k_1 = \tilde{k}_1(\tilde{\lambda}^2_{tu}) \equiv \omega_* \frac{(1-\theta_1)(1+\tilde{\lambda}^2_{tu})^2}{\left[2+(3-\theta_1)\tilde{\lambda}^2_{tu}\right]^2}.$$

This condition depends now on $\tilde{\lambda}^2_{tu}$ the endogenous multiplier associated to the budget-balance condition in country 2. The budget-balance condition in country 2 can be rewritten as follows

$$\omega_2 \frac{\tilde{\lambda}^2_{tu}(1+\tilde{\lambda}^2_{tu})}{(1+2\tilde{\lambda}^2_{tu})^2} + \omega_* \frac{(1+\tilde{\lambda}^2_{tu})(\tilde{\lambda}^2_{tu}+\theta_1)}{\left[2+(3-\theta_1)\tilde{\lambda}^2_{tu}\right]^2} = k_2.$$

It is immediate to notice that when $k_2 = \underline{k}_2 \equiv \frac{\omega_* \theta_1}{4}$ then $\tilde{\lambda}^2_{tu} = 0$ and $\tilde{k}_1(0) = \underline{k}_1$. From now on, we assume that k_2 is in a neighborhood of $\underline{k}_2$, which is equivalent to a small infrastructure deficit in country 2. Notice that this assumption is not necessary but simplifies dramatically the computations.

Moreover, this also implies that when $k_2 = \underline{k}_2$ if $k_1 = \underline{k}_1$ then country 1 does not need to subsidy the infrastructure and, as previously, we have $\tilde{\lambda}^1_{uu} = \lambda_{pf} = 0$.

Finally, when both countries adopt the users-pay system the budget-balance condition in country 2 can be rewritten as follows

$$\omega_2 \frac{\tilde{\lambda}^2_{uu}(1+\tilde{\lambda}^2_{uu})}{(1+2\tilde{\lambda}^2_{uu})^2} + \omega_* \frac{(1+\tilde{\lambda}^2_{uu})(\tilde{\lambda}^2_{uu}+\theta_1)(1+\tilde{\lambda}^1_{uu})^2}{\left[2+(3-\theta_1)\tilde{\lambda}^2_{uu}+\tilde{\lambda}^1_{uu}(2+3\tilde{\lambda}^2_{uu}+\theta_1)\right]^2} = k_2. \tag{21}$$

When k_1 goes to $\underline{k}_1$ and k_2 goes to $\underline{k}_2$, we already know that $\tilde{\lambda}^1_{uu}$ goes to 0. This entails that $\tilde{\lambda}^2_{uu}$ defined by (21) also goes to 0.

All these considerations will simplify the forthcoming computations. Indeed, totally differentiating (21) we get

$$\frac{\partial \tilde{\lambda}^2_{uu}}{\partial \tilde{\lambda}^1_{uu}}|_{k_1=\underline{k}_1} = \frac{\omega_* \theta_1^2}{4\omega_2 + (1-\theta_1)^2 \omega_*}. \tag{22}$$

When country 1 also adopts the users-pay system, the budget-balance condition can be rewritten as

$$\omega_1 \frac{\tilde{\lambda}^1_{uu}(1+\tilde{\lambda}^1_{uu})}{(1+2\tilde{\lambda}^1_{uu})^2} + \omega_* \frac{(1+\tilde{\lambda}^1_{uu})(1+\tilde{\lambda}^1_{uu}-\theta_1)(1+\tilde{\lambda}^2_{uu})^2}{\left[2+3(3-\theta_1)\tilde{\lambda}^2_{uu}+\tilde{\lambda}^1_{uu}(2+3\tilde{\lambda}^2_{uu}+\theta_1)\right]^2} = k_1. \tag{23}$$

Totally differentiating this condition, we obtain

$$\frac{\partial \tilde{\lambda}^1_{uu}}{\partial k_1}|_{k_1=\underline{k}_1} = \frac{1}{\omega_1 + \frac{\omega_*}{4}\left[\theta_1^2 - (1-\theta_1)^2 \frac{\partial \tilde{\lambda}^2_{uu}}{\partial \tilde{\lambda}^1_{uu}}|_{k_1=\underline{k}_1}\right]}, \tag{24}$$

$$= \frac{1}{\omega_1 + \frac{\omega_* \theta_1^2}{4}\left[1 - \frac{\omega_*(1-\theta_1)^2}{4\omega_2+\omega_*(1-\theta_1)^2}\right]}. \tag{25}$$

Given the previous considerations, and using (22) and (25), tedious but straightforward computations lead to

$$\frac{d\Delta SW_u^1}{dk_1}|_{k_1=\underline{k}_1} \propto -4\omega_2 + (1-\theta_1)\omega_*.$$

We conclude with the following lemma.

Lemma 3 *Assume that demands are linear, $c_d = 0$, $\lambda_{pf} = 0$ and that country 2 adopts the users-pay system. For k_2 sufficiently close to $\underline{k}_2$, in a neighborhood of $\underline{k}_1$ country 1 prefers to adopt the users-pay system than the taxpayers-pay one if $\omega_*\theta_2 - 4\omega_2 < 0$ and conversely.*

Finally, notice that whatever the parameters values we have $\underline{k}_1 \geq \tilde{k}_1(\tilde{\lambda}_{ut}^2)$. Our results can be extended to situation in which the infrastructure deficit in country 2 is not close to 0. In this case, the condition stated in Proposition 3 will be different and will depend on $\tilde{\lambda}_{ut}^2$ and $\tilde{\lambda}_{uu}^2$.

Numerical illustrations We propose three examples that illustrate the possible equilibria of the two-stage game. We always consider $\lambda_{pf} = 0$, $\alpha_i = 2$, $\gamma_i = \gamma_* = 1$, $c_u = 1$, α_* such that $\omega_* = 8\omega + \epsilon_1$, $k_i = \frac{\omega_*}{8} + \epsilon_2$ and $\theta_1 = \theta_2 = \frac{1}{2}$. Welfare are approximated to the fourth decimal.

- $\epsilon_1 = -\frac{1}{2}$ and $\epsilon_2 = 0.01$. The matrix of social welfare is (country 1 chooses the row, country 2 chooses the column)

	t	u
t	(0.8288,0.8288)	(0.825,0.83)
u	(0.83,0.825)	(0.8253,0.8253)

  In this first example, both infrastructure managers have the incentive to adopt the users-pay system whatever the choice in the other country. As a result the unique Nash equilibrium is (u, u).

- $\epsilon_1 = +\frac{1}{2}$ and $\epsilon_2 = 0.01$. The matrix of social welfare is (country 1 chooses the row, country 2 chooses the column)

	t	u
t	(1.1824,1.1824)	(1.1762,1.1844)
u	(1.1844,1.1762)	(1.1752,1.1752)

  In this second example, an infrastructure manager prefers to use the users-pay system when the other infrastructure manager adopts the taxpayers-pay system only. There exist then two pure-strategy asymmetric equilibria (u, t) and (t, u) and one mixed-strategy equilibrium.

- $\epsilon_1 = -2$ and $\epsilon_2 = 0.15$. The matrix of social welfare is (country 1 chooses the row, country 2 chooses the column)

	t	u
t	(0.3928,0.3928)	(0.3805,0.3818)
u	(0.3818,0.3850)	(0.3665,0.3665)

Finally, in this case, both infrastructure managers prefer to adopt the taxpayers-pay financing system and the unique Nash equilibrium is (t,t).

8.3.2 Strategic complementarity and iso-elastic demands

Let us assume that the domestic demand functions are given by $q_i(p_i) = p_i^{-\eta_i}$, $i = 1, 2$. The international demand is given by $q_*(p_*) = p_*^{-\eta_*}$. To have all the optimality conditions satisfied, we assume that the elasticity parameters are strictly greater than 1.

The methodology is identical and we content ourselves with the presentation of the main results.

Taxpayers-pay in country 2 Assume that country 1 adopts the users-pay system. If the fixed cost of the infrastructure in country 1 is such that

$$k_1 = \underline{k}_1 \equiv (1-\theta_1)\frac{(\eta_* - 1)^{\eta_* - 1}}{(2c_u\eta_*)^{\eta_*-1}\eta_*}, \tag{26}$$

then one can check that $\tilde{\lambda}^1_{ut} = 0$. Differentiating the budget-balance constraint in country 1 under a users-pay system, we obtain

$$\frac{\partial \tilde{\lambda}^1_{ut}}{\partial k_1}|_{k_1=\underline{k}_1} = \frac{1}{\frac{1}{\eta_1 c_u^{\eta_1 - 1}} + \frac{(\eta_*-1)^{\eta_*-1}\theta_1^2}{(2c_u\eta_*)^{\eta_*-1}\eta_*}}. \tag{27}$$

After tedious but straightforward computations, we find that

$$\frac{d\Delta SW_t^1}{dk_1}|_{k_1=\underline{k}_1} = \left[\frac{\partial \Delta SW_t^1}{\partial k_1} + \frac{\partial \Delta SW_t^1}{\partial \tilde{\lambda}^1_{ut}}\frac{\partial \tilde{\lambda}^1_{ut}}{\partial k_1}\right]|_{k_1=\underline{k}_1} \tag{28}$$

$$\propto \frac{\theta_1^2 c_u(\eta_* - 1)^{\eta_*}}{2^{\eta_*-1}(c_u\eta_*)^{\eta_*}(\eta_*-1)^2} > 0. \tag{29}$$

This enables us to state the following lemma.

Lemma 4 *Assume that demands are iso-elastic, $c_d = 0$, $\lambda_{pf} = 0$ and that country 2 adopts the taxpayers-pay system. In a neighborhood of $\underline{k}_1$ country 1 prefers to adopt the taxpayers-pay system than the users-pay one.*

Users-pay in country 2 First assume that country 1 adopts the taxpayers-pay financing system. Then one can check that when the fixed cost of the infrastructure in country 2 is such that

$$k_2 = \underline{k}_2 \equiv \theta_1 \frac{(\eta_* - 1)^{\eta_* - 1}}{(2c_u \eta_*)^{\eta_* - 1} \eta_*},$$

then $\tilde{\lambda}^2_{tu} = 0$. From now on, we assume that k_2 is in a neighborhood of $\underline{k}_2$. Moreover, when $k_1 = \underline{k}_1$ then country 1 becomes indifferent between choosing the taxpayers-pay or the users-pay system, implying that $\tilde{\lambda}^1_{uu} = 0$.

Differentiating the budget-balance condition in country 2 when both countries adopt the users-pay financing system we obtain

$$\frac{\partial \tilde{\lambda}^2_{uu}}{\partial \tilde{\lambda}^1_{uu}}|_{k_1 = \underline{k}_1} = \frac{2\eta_2 \theta_1^2 c_u^{\eta_2}}{2^{\eta_*}(\eta_* - 1)\left(\frac{c_u \eta_*}{\eta_* - 1}\right)^{\eta_*} + 2c_u^{\eta_2}\eta_2(1-\theta_1)^2}. \tag{30}$$

Differentiating the budget-balance condition in country 1 when both countries adopt the users-pay financing system and using (30) we obtain

$$\frac{\partial \tilde{\lambda}^1_{uu}}{\partial k_1}|_{k_1 = \underline{k}_1} = \frac{c_u^{\eta_1 - 1}\eta_1 \left[2^{\eta_*}(\eta_* - 1)\left(\frac{c_u \eta_*}{\eta_* - 1}\right)^{\eta_*} + 2\eta_2 c_u^{\eta_2}(1-\theta_1)^2\right]}{2^{\eta_*}(\eta_* - 1)\left(\frac{c_u \eta_*}{\eta_* - 1}\right)^{\eta_*} + 2\eta_2 c_u^{\eta_2}(1-\theta_1)^2 + 2\eta_1 c_u^{\eta_1}\theta_1^2}. \tag{31}$$

This enables us to obtain finally

$$\frac{d\Delta SW^1_u}{dk_1}|_{k_1 = \underline{k}_1} = \left(\frac{\partial \Delta SW^1_u}{\partial k_1} + \left[\frac{\partial \Delta SW^1_u}{\partial \tilde{\lambda}^1_{uu}} + \frac{\partial \Delta SW^1_u}{\partial \tilde{\lambda}^2_{uu}}\frac{\partial \tilde{\lambda}^2_{uu}}{\partial \tilde{\lambda}^1_{uu}}\right]\frac{\partial \tilde{\lambda}^1_{uu}}{\partial k_1}\right)|_{k_1 = \underline{k}_1} \tag{32}$$

$$\propto \frac{\frac{c_u}{2^{\eta_* - 1}}\theta_1^2 \left(\frac{\eta_* - 1}{c_u \eta_*}\right)^{\eta_*}\left[2^{\eta_*}(\eta_* - 1)\left(\frac{c_u \eta_*}{\eta_* - 1}\right)^{\eta_*} + 2c_u^{\eta_2}\eta_2\eta_*(1-\theta_1)\right]}{(\eta_* - 1)^2\left[2^{\eta_*}(\eta_* - 1)\left(\frac{c_u \eta_*}{\eta_* - 1}\right)^{\eta_*} + 2c_u^{\eta_2}\eta_2(1-\theta_1)^2\right]} > 0. \tag{33}$$

We conclude with the following lemma.

Lemma 5 *Assume that demands are iso-elastic, $c_d = 0$, $\lambda_{pf} = 0$ and that country 2 adopts the users-pay system. For k_2 sufficiently close to $\underline{k}_2$, in a neighborhood of $\underline{k}_1$ country 1 prefers to adopt the taxpayers-pay system than the users-pay one.*

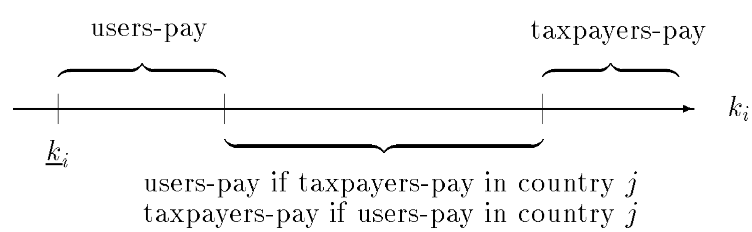


Figure 1: Country i's first stage choice of mode of regulation.

VERTICAL RELATIONSHIPS BETWEEN RAILWAY OPERATIONS AND THE INFRASTRUCTURE FOR THE EUROPEAN CASE *

Pedro Cantos Sánchez
Departamento de Análisis Económico
(Universidad de Valencia)
Edificio Departamental Oriental
Campus dels Tarongers, s/n
46022 Valencia (Spain)
Phone: 34 96 382 82 46
Fax:: 34 96 382 82 49
E-mail: Pedro.Cantos@uv.es

Abstract

A key question in the reform process of Europe's railway industry is the decision whether to opt for a market structure that vertically integrates infrastructure and operations, or for one that separates them. In making this decision, knowledge of the relationship between operating costs and the infrastructure is fundamental. For this purpose a translogarithmic cost function is estimated, including the multi-product nature of railway companies and a measurement of the value of the infrastructure. The results show that costs deriving from freight transport and from the infrastructure are complementary, while those deriving from passenger transport and from the infrastructure are substitutes. Therefore, if opting for a vertically separated structure the losses in efficiency of this process should be properly assessed.

Keywords: vertical integration, infrastructure, complementary costs, railways.

* The author wishes to thank R. Moner, L.Serrano and J. Campos for their comments and suggestions. I also wish to thank the participants in the Workshop on Infrastructure Charging on Railways held in Helsinki the 31st-July and 1st -August of 2000 for their comments, and the *Consellería de Cultura, Educación y Ciencia* de la *Generalitat Valenciana* (GV99-103-1-08) for financial support.

1. Introduction

There is an extensive bibliography on the testing for economies of scale in railway transport. An exhaustive and wide-ranging review of many of these studies appears in Oum et al. (1999). This concern is justified by the traditional hypothesis that this sector presented the conditions of a natural monopoly. On the other hand, the main result obtained from the first estimations for American companies (see Caves et al., 1981, 1982) was that economies of scale basically occur through more intensive use of the infrastructure rather than by expanding it. Thus economies of scale are associated with the use of the infrastructure and not with transport material. This result is explained by the fact that, for most studies, returns to scale (defined as the impact on cost of a proportional variation in traffic levels and of the network variable) are practically constant, whereas returns on density (defined as impact on cost of increasing traffic while maintaining the network size constant) are clearly increasing.

Estimations for European companies give similar results (see McGeehan, 1993, De Borger, 1992, Fillipini and Magi, 1993, Preston and Nash, 1996, Cantos and Maudos, 2000, and Cantos, 2000). The alternative approach, contrary to the traditional view, therefore proposes that totally differentiated bodies should manage infrastructure operations. Thus infrastructure (which fulfils the requisites of a natural monopoly) would be in the hands of one company, whether private or public, and one or more different firms would manage operations (which behave as a basically competitive or contestable market). The latter case offers the choice between granting a concession to one company to manage all operations (known as competition *for* the market) or a regime of free entry to operators (competition *in* the market). Recent directives of the European Commission seem to be aimed in this direction, in the sense that management and ownership of infrastructure and of services should be placed in the hands of different bodies.

Most of the processes of restructuring undertaken throughout the world have opted to maintain the vertical integration between infrastructure and operations, either through a system of competitive access in which trackage rights are offered (USA, Argentina, Japan or Brazil) or through a system of unified management but privatising the whole company (New Zealand). The European experiments, however, have opted for a scenario of vertical separation. Thus, in 1988, Sweden totally separated the ownership of infrastructure from services, yet keeping both bodies resulting from the reform in public hands. In 1994 the United Kingdom established a structure of vertical separation, and in 1996 began to privatise the infrastructure, while rail operations were offered to private enterprises under a regime of 25 franchises.

A description of the advantages and disadvantages of this process of vertical separation can be found in Campos and Cantos (2000). One of the advantages of this structure is that it places the railways in a similar situation to road transport in terms of tariff policy and infrastructure planning. Therefore, pricing (as pointed out by Nash, 1992) would be based on the criterion of social cost, and infrastructure planning would be based on cost-benefit analysis. Furthermore, the structure also facilitates the free entry process of operators, feasible on corridors or routes of high traffic density.

On the other hand there are important disadvantages that must also be properly evaluated. There are three principal disadvantages of this new structure for the sector: the loss of economies of scope deriving from the joint supply of infrastructure and operations, the risk that the new system may be less attractive to the users than an integrated system, and the reduction of incentives to investment by the company owning the infrastructure. Our study will evaluate the extent of the first disadvantage mentioned. It is often pointed out that the relationship between the services supplied and the rolling stock used, as well as the quality, quantity and technical characteristics is so close that both aspects need to be planned together. Thus, the assignment of different services to several operators may imply a lower utilisation of the staff and physical assets of the sector.

This study thus aims to evaluate these relationships between the costs of the European railways from 1973 to 1990. If the modifications of the structure significantly affect the costs of railway operations, it would be advisable to maintain a vertical structure in the industry, or at least the problems of co-ordination and inefficiencies deriving from the process of vertical separation should be compensated by the advantages of this process. This is a novel analysis that has not usually been dealt with in the literature, and a very important question in the definitive evaluation of a vertically separated structure.[1] To achieve this objective we consider a translogarithmic cost function with two levels of output (passengers and freight), and one variable representing the infrastructure, measured as the net value of ways and fixed installations. The interactions among these three

---

[1] Recently, Ivaldi and McCullough (2000) have found relevant vertical relationships between the infrastructure and the operations for a sample of US freight railways.

variables will enable us to analyse the relevance of the economies of scope between infrastructure and the different services, and between passenger and freight operations themselves.

The paper is structured as follows. Section 2 describes the data used in the study and Section 3 presents the model tested in the paper and the equations estimated. Section 4 presents the results while Section 5 ends with the conclusions and the principal recommendations deriving from this study.

2. The data.

The data used in the estimation were obtained from the reports published by the *Union Internationale des Chemins de Fer* (UIC) for the period 1973-1990.[2] We have disregarded some companies either by a lack of information (Irish company), or by the atypical character of the observations about some variables (Greek and Portuguese companies). There are 12 companies included in the model (see Table 1).

During the period of the estimation the market structure for all the companies was that of a public monopoly, though with varying degrees of autonomy from one country to another. The only exception is the Swedish company, *Statens Jarnvager* (SJ), which in 1989 separated into one firm owning and running the infrastructure, *Banverket* (BV), and one to

[2] From 1991 energy costs, purchases of materials and external services were aggregated into one single account, which did not allow the separation between the price of energy and of materials and external services to be maintained. Therefore we opt to reduce the sample to the period 1973-90.

run operations (SJ). As there are only two observations in this process, the conclusions that we could obtain are not significant.

The consideration of the multi-product nature of railway transport is a difficult problem to solve. Costs are very different for international, long distance or urban passenger transport. In the case of freight transport costs are also very different, distinguishing among general, intermodal or bulk traffic. Given the lack of more disaggregated information we opted, as is usual in the literature, to differentiate only between passenger traffic and freight traffic.

In particular, the variables referred to the *output* of the rail companies we use are the number of train-km of freight (Y_f) and of passengers (Y_p). As Nash points out (1985), these variables have some advantages over the usual specifications of the variables of *output* in transportation (ton-km. and passenger-km.). The volume of passengers-km and ton-km are directly influenced by the tax and subsidy policies established by the operators and their respective governments. Then they cannot be highly representative of the productive efficiency of the operating companies. However, the number of train-km represents a better approximation to the physical *output* offered.

As a measurement of the infrastructure we include in our model a monetary value of the ways and fixed installations held by the companies (I) as quasi-fixed input. This measurement is updated from the purchase or construction value of way and structures infrastructure net of depreciation. This variable also plays a key role in the determination of companies' costs, and reflects the changes in the value of the infrastructure, which is continuously being affected by the investments made by the companies. In fact, this variable may be considered as an additional input for the railways.

We also included a variable representative of the network length in our model , that is, the total number of track kilometres (L). This variable allows us to estimate the economies of scale, and it is different from the earlier variable, because I denotes the value of the total of infrastructures managed by the companies (buildings, ways, stations, etc.).

The operating costs (denoted by *TC*) have been taken as the dependent variable and include: i) labour costs, ii) energy and fuel costs and iii) material purchases and external services. We will consider three inputs or production factors. The labour costs divided by the total number of workers of the company are considered as an approximation to the price index of labour (w_1). We use the energy costs divided by the number of total train-km offered by each company as the price index of the energy inputs (w_2). Finally, we take the cost derived from materials and supplies divided by the number of each company's total train-km as the variable representative of the price of the materials and external supplies (w_3). This approximation to the *input* prices is similar to the one proposed by Preston and Nash (1996), given the impossibility of obtaining some more rigorous indices. All these variables were expressed in constant 1990 dollars, through the utilisation of the indices of the purchasing power parity (PPP) obtained from the available information in the OECD reports. Table 1 summarises the main statistics of the variables used in the sample.

(INSERT TABLE 1)

3. The estimation of the model.

In our case, all the operators in the period of estimation are public companies. Therefore the choice of the *output* levels is influenced by social considerations. From this point of view, it seems natural to suppose that the *output* levels are exogenously determined. If we further assume that the *input* prices are exogenous, then the rail companies that compose the sample will select that combination of *inputs* that minimise the production costs of the levels of *output* required.

A second order translog cost function will be estimated from the data previously presented. We will impose the standard conditions on the cost function in order to guarantee its good behaviour (linear homogeneity condition in input prices and symmetry in the estimated parameters). Applying the Sheppard lemma, we will obtain the participation equations in the costs for each factor. It is known that the joint estimation of the cost function and the participation equations increases the efficiency of the estimation without reducing the number of degrees of freedom. Since the three participation equations add up to unity, only two of them are linearly independent. For this reason, one of the demand equations will be eliminated. In any case, since we use the maximum likelihood estimator, the estimated parameters are independent of the omitted equation (Barten, 1967).

Dummy variables for each one of the individual companies are also included so that the estimators obtained are the *within-groups* estimators. This estimation allows one to control for the unobservable fixed effects for each one of the companies. There are variables which can be specific to each company, relatively stable along time and difficult to be introduced into the model. Some of these variables may be related to the network structure and the geographic characteristics of each railroad. We also include time effects in the model by introducing dummy variables for each year. The cost function and the cost share will be the following:

$$\begin{aligned} LnTC_t = {} & \alpha_0 + \sum_{j=p,f} \alpha_j \, LnY_{jt} + \frac{1}{2} \sum_{j=p,f} \sum_{k=p,f} \alpha_{jk} \, LnY_{jt} \, LnY_{kt} + \\ & + \sum_{i=1}^{3} \beta_i \, Ln\,w_{it} + \frac{1}{2} \sum_{i=1}^{3} \sum_{m=1}^{3} \beta_{im} Ln\,w_{it} \, Ln\,w_{mt} + \sum_{j=p,f} \sum_{i=1}^{3} \lambda_{ji} \, LnY_{jt} \, Ln\,w_{it} + \\ & + \sigma_I LnI_t + \frac{1}{2} \sigma_{II} \, LnI_t \, LnI_t + \sum_{j=p,f} \sigma_{Ij} \, LnI_t \, LnY_{jt} + \sum_{l=1}^{3} \sigma_{Ii} \, LnI_t \, Ln\,w_{it} + \\ & + \gamma_L \, LnL_t + \frac{1}{2} \gamma_{LL} \, LnL_t \, LnL_t + \sum_{j=p,f} \gamma_{Lj} \, LnL_t \, LnY_{jt} + \sum_{l=1}^{3} \gamma_{Li} \, LnL_t \, Ln\,w_{it} + \\ & + \sum_{s=1}^{11} X_s \, DUMC + \sum_{r=1}^{17} W_r \, DUMY + u_t \end{aligned} \qquad \textbf{(1)}$$

$$S_1 = \frac{\partial LnTC_t}{\partial Lnw_1} = \beta_1 + \sum_{i=1}^{3} \beta_{1i} \, Lnw_{it} + \sum_{j=p,f} \alpha_{1j} LnY_{jt} + \sigma_{I1} \, LnI_t + \gamma_{L1} \, LnL_t \qquad \textbf{(2)}$$

$$S_2 = \frac{\partial LnTC_t}{\partial Lnw_2} = \beta_2 + \sum_{i=1}^{3} \beta_{2i} \, Lnw_{it} + \sum_{j=p,f} \alpha_{2j} LnY_{jt} + \sigma_{I2} \, LnI_t + \gamma_{L2} \, LnL_t \qquad \textbf{(3)}$$

where t refers to the year, TC_{it} are costs, Y_{jt} is the vector of outputs, w_{it} is the vector of input prices, I_t represents the infrastructure variable, L_t denotes the length of the network, *DUMC* represents the dummy variables of each company, *DUMY* are the dummy variables for each year, and u_t the random disturbance term.[3]

The results of the estimation appear in table A.1. (see the Annex).[4] As can be observed, the parameters estimated have the expected signs. All the first order coefficients for input prices and for output are positive. As with previous results we can obtain estimates for economies of scale and density for the different companies of the sample. Note that returns with respect to traffic density are defined as the proportional variation of costs with proportional variation in levels of output, with the network size remaining constant. Meanwhile, returns to scale indicated are defined as the proportional variation of costs with proportional variation in levels of output and network size:

$$ED = \frac{1}{e_f^C + e_p^C} \qquad EE = \frac{1}{e_f^C + e_p^C + e_L^C}$$

where e^C_p, e^C_f and e^C_L measure, respectively, how the operating costs change proportionately due to a proportional variation in passenger traffic, freight traffic and the length of network. In

[3] In order to test if there are efficiency gains in the joint estimate of the cost function and the participation equations against the single estimate of the cost function, a maximum likelihood test was defined, indicating that the hypothesis of the joint estimation should be accepted. In particular, the value of the test was 1873,5, distributed as an X-squared with thirty-two freedom degrees.

[4] Given that some of the variables were not significant, a Wald test was defined in order to test the joint significance of these variables. The test result indicated that these variables should be maintained in the equation. The test value was of 79,99, distributed as an $F_{15,57}$.

other words, $\boldsymbol{e}^{C}_{p}$, $\boldsymbol{e}^{C}_{f}$ and $\boldsymbol{e}^{C}_{L}$ represent the cost elasticities of passenger traffic, freight traffic and the network size respectively. The results are shown in table 1. The results in terms of economies of scale are similar to those obtained recently for European companies (Preston and Nash 1996, Cantos 2000, SORT-IT, 1999). We thus observe that many companies present diseconomies of scale, and are therefore, too large. Regarding economies of density the results are more different with those obtained by the literature. In our case, the companies operate under increasing returns of traffic density, and therefore, a more intensive use of the infrastructure would enable a reduction of average costs.

(INSERT TABLE 2)

Table 3 shows the results obtained for the estimations of marginal costs per train-km of passengers and of freight. The marginal costs for passenger and freight traffic can be obtained from the respective expressions:

$$MC_p = \boldsymbol{e}^{C}_{p} \frac{TC}{Y_p}$$
$$MC_f = \boldsymbol{e}^{C}_{f} \frac{TC}{Y_f}$$

(INSERT TABLE 3)

The result that costs per train-km of freight are in general higher may be explained by the fact that the composition of freight trains are usually larger than passenger trains, and because the use of labour in freight transport is usually greater than in that of passengers (see Nash, 1985). We also observe that in general the companies with high passenger traffic densities have higher marginal costs freight traffic. The correlation coefficient obtained between the estimations for the marginal cost of passengers and the density levels of freight traffic (calculated as the quotient between the number of freight-km and the total km supplied by the companies) is 0.74. Moreover, companies that are more specialised in passenger traffic, such as the Dutch NS and the Danish DSB (see table 1), also have lower passenger marginal costs.

4. Vertical relationships.

We will analyse below the effects of variations in infrastructure and in output levels on different cost magnitudes, in particular on three parameters. Firstly, parameter a_{pf} can be expressed as follows:

$$a_{pf} = \frac{\partial LnTC_t}{\partial LnY_f \ LnY_p} = \frac{\dfrac{\partial LnTC_t}{\partial LnY_p}}{\partial LnY_f} = 0.1274 \quad (\textit{t-statistic} = 0.938)$$

Since the value of parameter a_{pf} is not significant at a level of 5% (though positive), this indicates that a proportional variation in freight transport will not significantly affect the cost elasticity of passenger transport. In other words, a proportional variation in the joint production of passengers and freight does not affect the proportion by which costs are altered.

This result indicates that the costs deriving from the two outputs are not complementary.[5] Therefore it does not seem that the separation of passenger and freight transport into two different companies will produce harmful effects on efficiency.

The most interesting comments are those referring to parameters s_{If} and s_{Ip}. These parameters can be written as follows:

$$s_{If} = \frac{\partial LnTC_t}{\partial LnI_t \, LnY_{ft}} = \frac{\dfrac{\partial LnTC_t}{\partial LnY_{ft}}}{\partial LnI_t} = -0{,}019 \qquad (t - statistic = -2{,}446)$$

$$s_{Ip} = \frac{\partial LnTC_t}{\partial LnI_t \, LnY_{pt}} = \frac{\dfrac{\partial LnTC_t}{\partial LnY_{pt}}}{\partial LnI_t} = 0{,}035 \qquad (t - statistic = 2{,}579)$$

Note that they are statistically significant at the level of 5% and have opposite sign. The first is negative, indicating that a proportional variation of infrastructure negatively affects the cost elasticity of freight transport. Thus the costs deriving from freight transport and from the infrastructure are complementary. On the other hand the positive sign of s_{Ip} indicates that a proportional variation of infrastructure positively affects the cost elasticity of passenger transport. Thus the costs of passenger output and of infrastructure are substitutes.

[5] The results previously obtained by the literature on this point are not conclusive (see, Cantos, 2000, and Preston and Nash, 1996).

In order to analyse these cost ratios in more detail, we estimated the elasticity of the marginal costs of each output relative to the other output. These elasticities appear in Table 4. Formally this elasticity is defined as:

$$\boldsymbol{e}_I^{MC_f} = \frac{\partial\, Ln\, MC_{ft}}{\partial\, Ln\, P_t}$$

$$\boldsymbol{e}_I^{MC_p} = \frac{\partial\, Ln\, MC_{pt}}{\partial\, Ln\, F_t}$$

(INSERT TABLE 4)

We thus find that, for the average of the sample, an increase of 1% in freight traffic increases the marginal cost of passenger transport by 0.07%, while an increase of 1% in passenger traffic increases the marginal cost of freight transport by 0.013%. In both cases, it should be remembered that these impacts are not statistically significant. These estimations, positive for all the companies estimated, reinforce the idea that the joint provision of passenger and freight transport has generated a situation of inefficiency.

Finally, we also calculate the elasticity of the marginal costs of each output relative to the infrastructure. Table 5 shows the results. Formally these elasticities are defined as:

$$\boldsymbol{e}_I^{MC_f} = \frac{\partial\, Ln\, MC_{ft}}{\partial\, Ln\, I_t}$$

$$e_I^{MC_p} = \frac{\partial \, Ln \, MC_{pt}}{\partial \, Ln \, I_t}$$

(INSERT TABLE 5)

The existence of complementarity of costs can be explained on the basis of a shared input that is not easy to divide and is important for the production of both outputs. In our case, the complementarity between freight transport and infrastructure may lie in the effect of, or need for, co-ordination between them. Freight transport usually takes place on certain groups of routes in large blocks or unit trains and normally requires complex routing and scheduling. In this case the additional endowment of infrastructure would reduce the costs of route designing and scheduling, and therefore reduce the cost of this traffic.[6] Therefore, decisions on infrastructure must take the needs of freight transport very much into account.

The substitute effect between infrastructure and passenger transport is more complicated to explain. Passenger traffic is usually dispersed over a large number of routes with very unequal traffic densities. In this case, an increase in infrastructure endowments would increase the marginal cost of transport per passenger, since the marginal increases in passenger traffic obtained are smaller than the cost magnitudes needed to produce them. Our results can be explained also because freight services may require more dedicated

[6] Ivaldi and McCullough (2000) have obtained a similar result for US freight railroads. They explain the complementarity between the infrastructure activity and general freight operations by means of this type of arguments.

infrastructure than passenger services and freight marginal costs would be more affected by the quality of infrastructure than passenger marginal costs.

Additionally we observe that the lowest size companies present greatest sensitivity of marginal cost to variations in the value of infrastructure, since they are the companies with greatest difficulties in obtaining new passengers from the increase in infrastructure.
In particular, the correlation coefficients between the estimations of ε^{MCp}_{I} and ε^{MCf}_{I} with the network size were -0.53 and -0.65 respectively. Therefore the need to co-ordinate infrastructure and transport of both passengers and freight also seems to be a question that must be taken into account.

5. Conclusions.

One of the key decisions regarding the future of the railway industry in Europe centres on the choice either of a structure of vertical integration that groups ownership of infrastructure and operations, or of a structure that separates ownership and management from each other. The choice of one structure or another depends on detailed analysis of the advantages and disadvantages of each (see Campos and Cantos, 2000). It is fundamental to analyse whether infrastructure costs and operating costs are related, i.e. if there are complementary or substitute effects between infrastructure and operations.

The results of this study indicate that the traffic costs of passengers and freight have been independent for the estimated period, 1973-90. We also observe that there are complementary effects between the costs deriving from freight transport and from infrastructure, while the effects between the costs deriving from passenger transport and the infrastructure are substitute. In any case, our results show that there are important vertical relationships between infrastructure and operations. Therefore, if important decisions regarding infrastructure are going to be made, rail-operating costs will be notably affected. Infrastructure and operations must be coordinated in order to maintain the coordination effects between both and to avoid possible inefficiencies. Obviously this result has been obtained for companies vertically integrated. But if these vertical relationships are present in a vertical unbundling structure, the risk of inefficiencies and loss of coordination effects between infrastructure and operations will be extremely high.

Thus if a vertical separation model is adopted rather than a vertical integration model (either with provision of infrastructure and operations by a single company, or with a regime of competitive access) the inefficiencies deriving from the resulting problems of co-ordination and from the loss of scope economies between infrastructure and operations must be properly evaluated. Finally, and independently of the estimated period, this paper has shown the relevance of the econometric techniques to analyse the existence of relationships among the different rail outputs and among the infrastructure and the rail operations. This question is fundamental in the future regulation of the rail market.

6. References.

Barten, A. (1967). "Maximum Likelihood Estimation of Complete Systems of Demand Equations". *European Economic Review,* May, pp.7-73.

Campos, J. and P. Cantos (2000). "Railways", in *Privatization and Regulation of Transport Infrastructure: Guidelines for Policymakers and Regulators*, ed. by De Rus, G. and A. Estache, World Bank Institute Development Studies.

Cantos, P. and J. Maudos (2000): "Efficiency, Technical Change and Productivity in the European Rail Sector: A Stochastic Frontier Approach". *International Journal of Transport Economics*, vol. 27(1), 55-75.

Cantos, P. (2000). "A Subadditivity Test for the Cost Function for the European Railways". *Transport Reviews*, 20(3), 275-290.

Caves, D. W., Christensen, L. R., and J. A. Swanson (1981). "Productivity Growth, Scale Economies and Capacity Utilization in US Railroads, 1955-1974". *American Economic Review*, vol. 71, December, pp. 994-1002.

Caves, D. W., Christensen, L. R., Tretheway, M. W. and R. J. Windle (1985). "Network Effects and the Measurement of Returns to Scale and Density in US Railroads". In Daugherty, eds., *Analytical Studies in Transport Economics*. Cambridge University Press.

De Borger, B. (1992). "Estimating a Multiple-Output Generalized Box-Cox Cost Function: Cost Structure and Productivity Growth in Belgian Railroad Operations, 1950-86". *European Economic Review*, 36, pp. 1379-1398.

Filippini, M. and R. Maggi (1993). "The Cost Structure of the Swiss Private Railways". *International Journal of Transport Economics,*19, 3, pp. 307-328.

Ivaldi, M. and G. J. McCullough (2000). "Density and Integration Effects on Class I U.S. Freight Railroads". *Journal of Regulatory Economics* (forthcoming).

McGeehan, H. (1993). "Railway Costs and Productivity Growth". *Journal of Transport Economics and Policy*, vol. 27, 1, pp. 19-32.

Nash, C. A. (1992). "The Rail Policy in the EC ", in De Rus, G. (coord.), *Economics and Transport Policy in Europe.* Ed. Civitas, pp. 97-123.

Preston, J. M., and C. A. Nash (1996). "The Rail Transport in Europe and the Future of RENFE", in José. A. Herce y G. de Rus (coord.), *The Transport Regulation in Europe*, Edit. Civitas, pp. 263-312.

OECD Statistics (2000). Purchasing Power Parities (available in www.oecd.org/statistics).

Oum, T. H. and C. Yu (1994). "Economic Efficiency of Railways and Implications for Public Policy". *Journal of Transport Economics and Policy*, 28 (2), 121-138.

Oum, T. H., W. G. Waters II and C. Yu (1999). "A Survey of Productivity and Efficiency Measurement in Rail Transport". *Journal of Transport Economics and Policy*, 33 (1), 9-42.

Table 1. Main Statistics

	L	*P*	*F*	*I*	*TC*	P_1	P_2	P_3	*%P*	*%F*
BR (UK.)	17389	340582	81192	1060300	5026920	15,8	0,68	4,06	0,81	0,19
DB (Germany)	28103	392979	196931	16538792	10073694	26,1	1,05	2,70	0,67	0,33
DSB (Denmark)	2253	40819	8189	624481	549220	15,5	0,97	3,34	0,83	0,17
FS (Italy)	16300	227361	58673	5787289	6518894	20,2	0,40	6,30	0,79	0,21
NS (Holland)	2845	98455	13311	1298291	825902	21,0	0,64	1,67	0,88	0,12
NSB (Norway)	4212	22811	10556	656399	332209	14,8	0,60	1,98	0,68	0,32
ÖBB (Austria)	5805	61791	36185	3677694	1904470	15,6	0,66	6,98	0,63	0,37
RENFE (Spain)	13084	100273	45412	3062479	1968894	19,8	1,16	3,72	0,69	0,31
SJ (Sweden)	11354	60424	40572	283524	914574	17,2	0,55	2,19	0,60	0,40
SNCB (Belgium)	4042	67959	22035	2013752	1932037	24,4	1,03	4,94	0,75	0,25
SNCF (France)	34900	291157	194560	9029953	7136700	20,0	0,59	4,13	0,60	0,40
VR (Finland)	5972	24832	18162	638826	462979	14,1	0,83	1,54	0,58	0,42

L: length of the network (in km).
P: km of passenger trains (in thousands).
F: km of freight trains (in thousands).
I: value of ways and fixed installations.
TC: total costs (in thousands of $).
P_1: labour price.
P_2: energy price.
P_3: price of materials and external services.
%P: percentage of the km of passenger trains with respect to the total km supplied.
%F: percentage of the km of freight trains with respect to the total km supplied.

Table 2. Economies of density and scale*				
	ED	Standard Error	EE	Standard Error
BR	1.48	0.020	0.53	0.008
DB	1.44	0.014	0.47	0.002
DSB	1.97	0.071	2.06	0.344
FS	1.45	0.001	0.65	0.041
NS	1.90	0.088	1.40	0.030
NSB	2.04	0.040	1.00*	0.014
ÖBB	1.73	0.073	0.85	0.007
RENFE	1.60	0.015	0.60	0.005
SJ-BV	1.87	0.064	0.63	0.003
SNCB	1.83	0.067	1.08*	0.080
SNCF	1.42	0.030	0.45	0.003
VR	2.02	0.039	0.96	0.007
Average	1.73	0.238	0.83	0.473

* Results non-statistically different from 1 at 5% (the rest of the values are statistically different from 1 at 5%).
ED= Economies of density. EE= Economies of scale.

Table 3. Marginal costs for output activities				
	MC_f	Standard Error	MC_p	Standard Error
BR	15.06	6.81	6.21	1.57
DB	6.63	1.34	13.04	3.53
DSB	20.99	4.60	3.89	0.76
FS	23.37	8.86	12.93	4.82
NS	19.15	7.49	2.61	1.01
NSB	9.44	3.08	3.69	1.37
ÖBB	11.11	3.41	12.38	4.21
RENFE	9.07	2.62	7.75	1.89
SJ-BV	6.01	2.07	4.06	1.74
SNCB	22.75	5.89	10.13	3.92
SNCF	6.41	11.88	11.88	3.16
VR	6.08	1.84	4.45	1.69
Average	12.94	8.23	7.86	4.86

Table 4. Elasticities of marginal costs with respect to output				
	e^{MCp}_{f}	Standard Error	e^{MCf}_{p}	Standard Error
BR	0.048	0.010	0.067	0.012
DB	-0.048	0.008	0.191	0.017
DSB	0.151	0.010	-0.079	0.020
FS	0.033	0.010	0.104	0.023
NS	0.139	0.017	-0.064*	0.037
NSB	0.146	0.010	-0.119	0.017
ÖBB	0.042	0.010	0.050*	0.033
RENFE	0.037	0.006	0.050	0.010
SJ-BV	0.111	0.015	-0.100	0.030
SNCB	0.093	0.025	-0.017*	0.040
SNCF	-0.033	0.013	0.162	0.033
VR	0.113	0.013	-0.103	0.013
Average	0.068*	0.066	0.013*	0.108

All the values are statistically significant at 5%, except the values denoted by *.

Table 5. Elasticities of marginal costs with respect to infrastructure				
	e^{MCf}_{I}	Standard Error	e^{MCp}_{I}	Standard Error
BR	-0.052	0.008	0.119	0.006
DB	-0.143	0.012	0.076	0.005
DSB	-0.053	0.005	0.132	0.011
FS	-0.081	0.008	0.905	0.010
NS	-0.027	0.008	0.156	0.020
NSB	-0.076	0.005	0.133	0.013
ÖBB	-0.116	0.008	0.063	0.013
RENFE	-0.099	0.004	0.082	0.004
SJ-BV	-0.065	0.008	0.145	0.020
SNCB	-0.073	0.014	0.106	0.021
SNCF	-0.138	0.013	0.070	0.007
VR	-0.091	0.006	0.118	0.010
Average	-0.085	0.035	0.108	0.032

(All the values are statistically significant at 5%).

ANNEX.

Table A.1. Estimate results

	Parameter	Standard Error	t-statistic
α_0	14,072	0,517	27,180
β_1	0,678	0,007	93,659
β_2	0,077	0,002	33,043
α_p	0,411	0,068	5,985
α_f	0,204	0,092	2,227
σ_I	-0.066	0,016	-0,401
γ_L	1,253	0,307	4,082
β_{11}	0,131	0,011	11,569
β_{12}	-0,023	0,003	-8,622
β_{22}	0,043	0,002	19,838
β_{23}	-0,033	0,013	-2,542
λ_{f1}	0,015	0,017	0,859
λ_{f2}	-0,012	0,007	-1,694
λ_{p1}	-0,046	0,010	-4,235
λ_{p2}	0,012	0,003	4,242
σ_{I1}	0,114	0,005	2,422
σ_{I2}	-0,002	0,001	-1,731
γ_{L1}	0,024	0,017	1,398
γ_{L2}	0,001	0,007	0,253
α_{pp}	0,038	0,142	0,270
α_{ff}	-0,048	0,254	-0,191
σ_{II}	-0,009	0,013	-0,673
γ_{LL}	0,726	0,310	2,338
σ_{If}	-0,019	0,043	-2,446
α_{pf}	0,012	0,092	0,938
σ_{Ip}	0,035	0,022	2,579
γ_{LI}	-0,022	0,037	-0,600
γ_{Lp}	-0,038	0,117	-0,328
γ_{Lf}	0,099	0,243	0,409
X1	1,263	0,339	3,720
X2	0,522	0,099	5,235
X3	1,997	0,781	2,555
X4	1,529	0,371	4,112
X5	1,970	0,758	2,598
X6	1,666	0,742	2,244
X7	2,263	0,673	3,361
X8	1,251	0,468	2,667
X9	1,064	0,522	2,039
X10	2,302	0,738	3,117
X12	1,709	0,684	2,498
W74	0,021	0,016	1,297
W75	0,040	0,020	1,971
W76	0,036	0,020	1,798
W77	0,032	0,020	1,605
W78	0,039	0,020	1,917

W79	0,033	0,020	1,685
W80	0,040	0,019	2,137
W81	0,054	0,020	2,685
W82	0,049	0,021	2,335
W83	0,046	0,020	2,329
W84	0,038	0,023	1,623
W85	0,042	0,016	2,490
W86	0,033	0,016	2,042
W87	0,011	0,016	0,687
W88	-0,021	0,017	-1,230
W89	-0,018	0,014	-1,234

Log of Likelihood Function = 1443,91
Number of Observations = 204
R^2 = 0,998

Equation S_1: R^2 = 0,697
Equation S_2: R^2 = 0,77

Essential Facility Financing and Market Structure*

Bernard Caillaud[†] and Jean Tirole[‡]

This version: December 12, 2000

Abstract

The paper analyzes the funding of an infrastructure project (high speed train line, platform, tunnel, harbor, regional airport, fiber-to-the-home network,...) in a situation in which an incumbent operator has private information about market profitability (demand, cost) and the infrastructure owner is subject to a budget constraint, either on a per project basis or over the entire infrastructure. An open access policy raises welfare, but may make the project non-viable since funding must be provided by capital contributions and access charges.

The infrastructure owner can ask the incumbent for a higher capital contribution if the latter insists on an exclusive use. Yet, such screening is at odds with social goals: The incumbent is willing to pay more for exclusivity, the higher the demand (the lower the cost), that is precisely when competition yields the highest benefits. At the optimum, the incumbent's information impacts the decision of whether to build the infrastructure, but is not used to determine market structure.

The paper further shows that an absence of long-term licencing favors monopoly franchising, while a threat of regulatory capture creates an open-access presumption.

*This paper is part of a joint research program between CERAS-ENPC and IDEI on the economics and the regulation of railroads. Financial and intellectual support from Réseau Ferré de France and the Direction des Transports Terrestres (French Ministry of Transportation, PREDIT Program) are gratefully acknowledged. We thank the participants to the June 8-9, 2000 conference on "New Developments in Railroad Economics: Infrastructure Investment and Access Policies" (Paris), the "Railraod Conference" (Helsinki) and the economic theory seminar at Tinbergen Institute, Jerry Hausman, Robert Willig and, especially, Antonio Estache and Jean-Jacques Laffont for helpful discussions and comments.

[†] CERAS-ENPC (URA 2036, CNRS), Paris and CEPR, London.

[‡] CERAS-ENPC (URA 2036, CNRS), Paris, IDEI, Toulouse and MIT, Cambridge.

1 Introduction

Scope of the analysis

The last twenty years have witnessed the large scale deregulation of sectors such as telecommunications, transportation and energy.[1] Activities with no significant returns to scale have been opened to competition and are now subject only to light regulation. In contrast, the other, "infrastructure" activities with significant returns to scale, large sunk costs and network externalities (e.g. cable, narrow-and-broadband copper local loops, transmission grids, harbors, regional airports, pipelines, intermodal platforms, high-speed train lines, Eurotunnel), are deemed to be essential facilities and are often regulated as public utilities or awarded exclusive franchising contracts.[2]

This paper studies investment and funding of an infrastructure project when the infrastructure is subject to a budget constraint either on a per-project basis or over the entire infrastructure. For example, a regulated railroad infrastructure owner may be required to invest only in projects that generate enough maintenance cost savings or raise enough access revenues to be financially viable,[3] or else to break even over its entire network. The downstream segment is populated by one incumbent and one (or several) entrant. In our model, the incumbent operator and the infrastructure owner will be taken to be separate entities, but they can indifferently be a single actor, which allows for a wide range of applications. For example, while incumbent railroad operators in Europe (respectively, electricity generators in the US) have been forced to divest the essential facilities, they are still vertically integrated in the US (respectively, in Europe).

The downstream market is potentially competitive, but the regulator can cut special deals to limit competition if that is what it takes to make the infrastructure financially viable.[4] Since investment profitability is driven by expected future revenues from charging

[1] See, e.g., Armstrong-Cowan-Vickers 1994, and Laffont-Tirole 1999.

[2] The reader will find useful material on franchises in the transportation sector in several World Bank reports (Campos-Cantos (1999), Estache (1999) and Kerf *et al* (1998)).

[3] As is the case in France for RFF, the infrastructure owner (Article 4, Order 97-444).

[4] To think about the link between downstream market design and upstream investment financing, two examples related to railways are worth keeping in mind. First, most of the new tracks that have been installed lately in France are high-speed tracks, that meet mechanical and safety standards to allow the

for access, the incentives to invest in infrastructure are related to market design at the downstream level. The static vision of access policies is thus put into a new perspective, as possibly resulting from an ex ante optimal arrangement to finance, build and operate infrastructure projects.[5]

Although our analysis is primarily motivated by infrastructure investment in regulated sectors (transportation, energy, telecommunications), it also sheds light on the twin issue of the treatment of essential facilities in antitrust law. Essential facilities have been defined by American and European competition authorities as facilities the access to which is essential (and not just cheaper than the alternative) in order to compete on the downstream market, and whose owner is dominant and has no valid reason (lack of capacity, cost of achieving interoperability, protection of IP rights, ...) to deny access. Competition authorities often face a dilemma between granting generous access to the essential facility and thereby generating substantial social benefits from competition, and letting the facility owner recoup its investment. Indeed, competition authorities usually exclude patented innovations from the scope of application of the essential facility doctrine even though such innovations often meet the essential facility criteria, on the grounds that the patentees' freedom to choose their licensees and licensing fees encourages innovation. Competition authorities face "ex post" the same dilemma as the regulators in regulated industries do "ex ante": More competition increases welfare once the facility is built, but

French high-speed train TGV to run at maximum efficiency. These tracks are specifically reserved to TGV traffic, although ordinary long-distance trains could run on them as well and compete with the TGV. Furthermore, the technical specifications of railbeds were selected so as to be incompatible with the German high-speed train. Such decisions involve a monopolistic market design.

The second example is Eurotunnel. Financing the tunnel was complex, and negotiations ended up with the following arrangement for access. Half of the slots were reserved to the Shuttle (operated by Eurotunnel), the other half was allocated to the British and French incumbent operators. This controversial policy was relaxed by European competition authorities, which imposed the release of slots for potential entrants.

We do not pretend that our analysis rationalizes or invalidates these decisions, all the more as these policies date back to the pre-liberalization epoch. It proposes, however, a basic formulation of the potential trade-offs that can help us think about future infrastructure projects.

[5] Along similar lines, Campos and Cantos (1999) argue that the allocation of exclusivity rights in mixed concession contracts (as in Argentina or Burkina-Faso and Côte d'Ivoire) reflects a balance between the benefits of allowing competitive access and the private operator's greater profit stream under a monopoly regime.

may result in a lower-quality facility or in no facility at all.

Eliciting project viability

We assume that the regulator has imperfect information about the future demand for (or future cost of providing) the services enabled by the infrastructure. Typically, the demand for a new platform, a tunnel, a highway or fiber-to-the-home bandwidth is hard to forecast. In this respect, operating companies', in particular incumbent ones', superior marketing expertise and production experience put them in a better position to assess the value of infrastructure investments. This observation suggests that the regulator should elicit the information about infrastructure viability held by the downstream industry. In this paper, we are mainly interested in situations in which one incumbent operator has private information about the demand for (or the cost of operating) the segment; we will also briefly discuss the possibility that entrants also have private information about the segment's profitability.

Because competition destroys profits, a restrictive access policy granting exclusive access to the incumbent generates higher financial returns than a more liberal one that opens access to entrants. The regulator is therefore more likely to extract high access charges for exclusive access, which helps finance the project. On the other hand, social welfare is maximized by a more competitive market structure. A fully informed regulator would implement the project under open access provided that competition is compatible with access charges that are high enough to balance the budget; otherwise, the regulator would grant a monopoly franchise to the incumbent operator, or even not invest at all.

When the incumbent operator has superior information about the value of the project, the regulator must elicit this information in an incentive compatible way. The screening instrument is the stake taken by the incumbent. The incumbent operator must contribute in order to have the new infrastructure built; furthermore, it must participate more to the financing if it is to demand exclusivity. Alas, private incentives conflict with the socially desirable policy of opening the market to competition when demand is high; for, it is precisely when demand is high that the incumbent is most eager to preserve its monopoly position. That is, the incumbent is more eager to bear a high fraction of the investment cost to secure an exclusive right when demand is high. The reader familiar with incentive theory will recognize here a situation of "non-responsiveness,"[6] in which incentive compatibility imposes a monotonicity of policy with respect to the agent's private information that is opposite to the one desired by the principal.

The policy implication of this theoretical result is that the infrastructure owner should not try to screen the incumbent operator by demanding a basic open access contribution with the option of a higher investment contribution in exchange of an exclusive right. Rather, the infrastructure owner should conduct its own in-depth marketing studies; it should also encourage the acquisition by potential entrants of information about demand, so that they can challenge undue claims made by the incumbent that demand is low and project viability requires an exclusive franchise.

Outline

The paper is organized as follows. Section 2 lays out a simple model of infrastructure financing with a per-project break-even constraint. Section 3 analyzes the perfect information benchmark. Section 4 identifies the fundamental departure from the benchmark induced by asymmetric information and derives the optimal policy. Extensions of the model are explored in the following sections.

In Section 5, the regulator is allowed to break even over the entire network rather than on a per-project basis. Section 5 shows that, whether the incumbent is the same or differs across projects, open access segments should be cross subsidized by monopoly franchises. It is still the case, though, that due to incomplete information, the most profitable segments are operated under a monopoly franchise. Section 6 investigates dynamics and shows that monopoly franchises are immune to the ratchet effect, while the incumbent has an incentive to signal a low profitability of the segment by turning down an open access license. Section 7 shows that a concern about regulatory capture may make open access the default rule and thus put the burden of proof on the monopoly

[6] See Guesnerie-Laffont (1984) and Caillaud et al (1988).

franchising option. Section 8 investigates two further extensions of the model, associated with the possibility of replacing the incumbent, and the observability of profits. Section 9 concludes.

Relationship to the literature

This paper belongs to a small, but growing literature on endogenous market structures. In Auriol-Laffont (1992) and Dana-Spier (1994), a regulator trades off the costs of duplication with the benefits of competition in a private value environment that is, in a situation in which operators have private information about their own production costs. Our research differs from theirs in two respects. First, the cost of competition is in terms of the viability of financing rather than duplication. Second, we consider a common value environment where the incumbent has private information about industry-wide parameters such as demand or cost. This departure turns out not to be innocuous: in a private value environment, the regulator is more willing to grant an exclusive franchise to an efficient firm, and conversely an efficient firm is willing to pay more than an inefficient one for the right to be the sole producer. Thus, the non-responsiveness that is central to our paper does not arise in Auriol-Laffont and Dana-Spier. Endogenous market structure determination also arises in the literature on universal service obligations (Milgrom 1996, Laffont-Tirole 1999), but this literature also focuses on private values and the set of issues considered there is rather different from those analyzed here.

2 Model

Consider an industry characterized by vertical separation between an upstream infrastructure and downstream service provision. The regulator, who in this version is also the upstream monopoly provider of infrastructure services, has to decide whether to build a new piece of infrastructure / realize a new project (equivalently, it could decide whether to keep an existing infrastructure in operation). The project involves investment cost I. Two operators can provide services over the infrastructure. The incumbent operator, firm 1, is already active in this market (or related segments, in the case of a new infrastructure) and has private information about the demand for and/or the cost of providing services. A potential entrant, firm 2, can enter the market and provide competitive services. The potential entrant, however, does not have access to the incumbent's information on market profitability before entry; it learns all relevant information about demand and cost only after it has entered and started operating services.[7] The regulator chooses whether to invest, and, in case of investment, whether the incumbent keeps a monopoly situation or the entrant is allowed to enter.

A parameter $\theta \in [\theta_L, \theta_H]$ characterizes the value of the project, both with respect to private profitability and to social welfare. To fix ideas, we will interpret it to be a downstream demand parameter, but it could also summarize common value aspects of the cost of providing services over the infrastructure. If the infrastructure is built, $\pi_M(\theta)$, $\pi_1(\theta)$ and $\pi_2(\theta)$ denote the (expected) profits made respectively, by an incumbent monopolist, and by the incumbent and the entrant in a duopoly. Similarly, let $W_M(\theta)$ and $W_D(\theta)$ denote social welfare, i.e. the sum of operators' profits and consumers' net surplus ($S_M(\theta)$ and $S_D(\theta)$) under monopoly and duopoly, respectively.

Let us spell out the information structure in more detail:

1. The incumbent operator knows θ. The regulator and the potential entrant do not; they share prior beliefs summarized by an absolutely continuous c.d.f. $F(.)$, with density $f(.)$. This assumption, while extreme, reflects the informational advantage that historical operators have over potential competitors and regulatory authorities. It could be relaxed by allowing the entrant to have some independent private information about market profitability.[8]

2. Profits (or social welfare levels) are not observable so that profit-based regulatory contracts, such as profit-sharing or profit-contingent fees, are not feasible. This assumption is motivated by the possibility of transfers across product lines or by

[7] In Section 6.2 we look at a two-period situation where the entrant learns by operating.

[8] Private information shared by the incumbent and the entrant could be costlessly elicited using a Maskin-type mechanism (Maskin (1999)); hence, only independent information induces inefficiencies of the type described in our paper.

moral hazard considerations: see section 8.2.

3. Price regulation is ruled out. This restriction can also be motivated by moral hazard considerations. For example, quality aspects of services cannot be perfectly assessed by regulatory authorities. The fine structure of prices under price discrimination for various categories of consumers might also be difficult to regulate. We could of course allow price regulation on some basic services (as is done for example for railroad franchises in the UK) as long as the incumbent has private information about the cost of the demand for these services or enhanced services.

We make the following assumptions:

A.1: The surplus and profit functions are increasing in the demand parameter θ.

A.2: For all θ, $\pi_M(\theta) \geq \pi_1(\theta) + \pi_2(\theta)$: competition reduces industry profit. To avoid corner solutions, we will further assume that: $\pi_M(\theta_L) < I < \pi_1(\theta_H) + \pi_2(\theta_H)$.

A.3: For all θ, $0 < W_M(\theta) \equiv S_M(\theta) + \pi_M(\theta) < W_D(\theta) \equiv S_D(\theta) + \pi_1(\theta) + \pi_2(\theta)$: social welfare is larger under duopoly than under monopoly.

A.4: For all θ, $\dot{\pi}_M(\theta) > \dot{\pi}_1(\theta)$: a monopolist captures a higher share from an increase in demand (as measured by an increase in θ) than the same firm operating as a duopolist.

Assumptions A.2 (competition reduces industry profit) and A.3 (competition increases welfare) are standard in industrial organization. If the services supplied by both operators are perfect or close substitutes, assumption A.2 is satisfied by standard models of imperfect competition. If, however, both operators propose significantly differentiated services, or if there were complementarities between the two operators, assumption A.2 might not hold, but presumably open access would then not be contentious. Assumption A.3 is likely to hold as long as there are limited returns to scale at the downstream level: then, increasing the number of operators does not involve production inefficiencies, e.g., fixed cost duplication, and more competition is unambiguously socially preferable. Again, open access is an issue only if there are benefits to competition. Assumption A.4, a sorting or Spence-Mirrlees condition, is central to the incentive analysis. It is quite natural, provided that competition is not profoundly altered when demand conditions, as measured by θ, change. For example, this set of assumptions is satisfied when the firms wage Bertrand-like price competition or when demand is linear with intercept θ and Cournot competition prevails after entry.[9]

The regulator maximizes aggregate social welfare, that is the sum of consumers' net surplus and producers' profits, under a balanced budget constraint. The regulator has two instruments: the market structure and investment contributions / access charges paid by the operators. First, the regulator can either grant a monopoly franchise to the incumbent, or give access to the infrastructure to both the incumbent and the entrant (or simply not build the infrastructure). In a first step, we rule out the possibility that the regulator grants a monopoly franchise to the entrant. We relax this assumption later on. Second, the regulator specifies the access charges paid by operating companies to use the infrastructure. In this model there is no difference between a contribution to the investment outlay and an access charge. Let a_M, a_1 and a_2 denote the access charges paid by firm 1 when it is a monopoly, and by firm 1 and firm 2 in case of a duopoly, respectively.

The financial constraint that forbids to invest without securing sufficient resources to match the investment spending can take different forms. A strong budget constraint requires that any project, characterized by its own θ, be fully financed by access charges: we call this an ex post budget balance constraint. A weaker constraint simply requires that the regulator balances the budget on average over all projects, so that the budget constraint is considered ex ante. Sections 3 and 4 consider the strong form, under which the regulator secures the funding for the individual project at the investment date. Section 5 analyzes the weaker form.

[9] Normalizing operating costs to zero and assuming perfect information about θ after entry, profit and surplus functions are given by $\pi_M(\theta) = \theta^2/4$ and $S_M(\theta) = \theta^2/8$, under monopoly, and by $\pi_1(\theta) = \pi_2(\theta) = \theta^2/9$ and $S_D(\theta) = 2\theta^2/9$ after entry.

Note that we ignore the contract renegotiation concerns that are so pervasive for instance for fixed-price concession contracts.[10] While the parameter θ represents the best forecast of demand at the investment date, the realized demand is affected by ulterior shocks (thus, our profit and welfare functions, which are contingent on θ, are already expectations taken over post-investment realizations of uncertainty). Although our theory could be generalized to allow for future renegotiations (in which case the functions $\pi_{\cdot}(\cdot)$ and $W_{\cdot}(\cdot)$ would embody the future renegotiation), the model as it stands assumes that the payment of the access charges is made credible by the posting of bonds or else that access charges take the form of upfront investment contributions granting future usage rights to the operators.

Remark: As announced in the introduction, the model applies without any modification to the vertically integrated case, in which the incumbent operator owns the infrastructure (the regulator is then necessarily a separate entity). The only difference relates to accounting: The incumbent receives $\pi_M(\theta) - I$ in case of monopoly franchise, and $\pi_1(\theta) + a_2 - I$ in case of open access.

3 Perfect information benchmark: the conflict between competition and viability

Suppose first that demand is commonly known.[11] The regulator is subject to the strong balanced-budget requirement and must then compare three policies:

- Not building the infrastructure yields no additional welfare.
- Building the infrastructure and granting a monopoly franchise to the incumbent with access charge $a_M(\theta)$ yields additional welfare equal to $W_M(\theta)$ provided budget is balanced: $a_M(\theta) \geq I$, and the incumbent breaks even: $\pi_M(\theta) - a_M(\theta) \geq 0$. In this

[10] See e.g. Engel et al (1997, 1998).

[11] We maintain throughout the assumption that profits and prices are unobservable, hence cannot be regulated.

case, it is clear that any access charge within $[I, \pi_M(\theta)]$ is feasible, provided that:

$$\pi_M(\theta) \geq I \tag{1}$$

that is, provided that:

$$\theta \geq \theta_M \equiv \pi_M^{-1}(I) \tag{2}$$

where $\theta_L < \theta_M < \theta_H$ from assumption A.2.

- Building the infrastructure and allowing competition with access charges $a_1(\theta)$ and $a_2(\theta)$ yields additional welfare equal to $W_D(\theta)$ provided the regulator balances her budget: $a_1(\theta) + a_2(\theta) \geq I$, and the two operating companies break even: $\pi_i(\theta) \geq a_i(\theta)$ for $i = 1, 2$. Again, many pairs of access charges are feasible provided that:

$$\pi_1(\theta) + \pi_2(\theta) \geq I \tag{3}$$

that is, provided that:

$$\theta \geq \theta_D^{SI} \equiv (\pi_1 + \pi_2)^{-1}(I) \tag{4}$$

where "SI" stands for "Symmetric Information", and where $\theta_M < \theta_D^{SI} < \theta_H$ from assumption A.2.

We have assumed that social welfare is higher under a duopolistic market structure. The infrastructure income generated by a duopolistic structure may, however, not be sufficient to finance the investment. Because competition reduces industry profit, the budget balance requirement may force the regulator to grant a monopoly franchise to the incumbent operator, a second-best solution. This is summarized in the following and very intuitive proposition.

Proposition 1 *: The optimal full-information investment and market design are given by:*

- *for $\theta \in [\theta_L, \theta_M)$, the infrastructure is not built,*

- *for $\theta \in (\theta_M, \theta_D^{SI})$, the infrastructure is built and a monopoly franchise is granted to the incumbent,*
- *for $\theta \in (\theta_D^{SI}, \theta_H]$, the infrastructure is built and operated under open access.*

An infrastructure project that generates competitive operating profits in excess of the investment cost should be implemented under open access. We noted that in this case, the access fees charged to the incumbent and to the entrant are indeterminate. There may be scope for negotiation about the relative magnitude of these charges. The investment decision and the market structure in contrast are unambiguous.

Open access however, jeopardizes operating profits. In the absence of adequate financing by customers, or other stakeholders who benefit from the project, a monopoly franchise may have to be granted to the incumbent operator, whose operating profit can then be partially captured by the regulator to cover the investment cost.

4 Incomplete information

We now turn to the more realistic framework where the incumbent operator, due to his experience in serving the market or similar markets, has private information about the demand (or cost) parameter θ, while the regulator and the entrant only have prior distribution $F(.)$ over this parameter.

Following the standard mechanism design approach, the regulator proposes a menu of decisions about investment, market design and access charges with ex post balanced budget, among which the incumbent chooses one. Decisions, in particular access charges, must guarantee the potential entrant non-negative net expected profits.[12] The mechanism can w.l.o.g. be restricted to be a revelation mechanism. The investment and market design decisions can take value within $\{0, M, D\}$,[13] where we let M denote the decision to build

[12] So, we assume that the entrant has to apply for access ex ante, before the mechanism takes place. Later, we show that the optimal (deterministic) mechanism is valid even if the entrant applies for entry after the incumbent's choice. So our assumption involves no loss of generality.

[13] So, we rule out stochastic mechanisms for the moment; see more on this below and in the appendix.

the project under franchised monopoly, D the decision to build the project under open access and 0, the decision not to build the project. Let $a_I(.)$ and $a_E(.)$ denote the charges paid by the incumbent and the entrant, conditional on the incumbent's announcement.

4.1 Incentive-compatible franchises

The technical analysis that follows is standard (see e.g. Guesnerie-Laffont [1984]). The first lemma shows that incentive compatibility severely constrains feasible allocations:

Lemma 2 *: For any (deterministic) incentive compatible mechanism, there exist at most two threshold levels, θ_* and θ^*, with $\theta_L \leq \theta_* \leq \theta^* \leq \theta_H$, such that:*

- *if $\theta \in [\theta_L, \theta_*)$, the infrastructure is not built;*
- *if $\theta \in (\theta_*, \theta^*)$, the infrastructure is built and operated under open access;*
- *if $\theta \in (\theta^*, \theta_H]$, the infrastructure is built and operated under a monopoly franchise.*

Proof. Suppose that for θ the infrastructure is built under a monopoly franchise while for $\theta' > \theta$, it is built under open access. Then incentive compatibility requires that:

$$\begin{aligned}\pi_M(\theta) - a_I(\theta) &\geq \pi_1(\theta) - a_I(\theta')\\ \pi_1(\theta') - a_I(\theta') &\geq \pi_M(\theta') - a_I(\theta),\end{aligned}$$

Summing up both inequalities yields:

$$\pi_1(\theta') - \pi_1(\theta) \geq \pi_M(\theta') - \pi_M(\theta), \tag{5}$$

which contradicts A.4. The comparisons with the decision not to build follows similar lines and relies on A.1. ■

When the incumbent has private knowledge about the profitability of the project, there is no way to make him forego a monopoly position as its information about profitability improves. If the regulator is ready to grant a monopoly franchise for some θ, she must also grant the monopoly franchise for higher values $\theta' > \theta$. So, the full information policy is not incentive compatible.

As is standard in the theory of mechanism design, once the decision function is fixed, the access charge $a_I(.)$ for the incumbent is completely determined up to a constant. This access charge is necessarily a step function, characterized by (a_0, a_1, a_M), depending on whether the decision is $0, D$ or M. Obviously $a_0 = 0$.

Conditionally on the infrastructure being built, the only menu the regulator can offer consists of two options. If the incumbent selects a "base contribution" a_1, then the infrastructure is built and access is opened to competition. If in contrast, the incumbent is willing to contribute $a_M > a_1$, then he is awarded an exclusive right. Moreover, given the previous lemma, the incumbent chooses the base contribution for $\theta \in (\theta_*, \theta^*)$, while he opts for exclusivity for $\theta > \theta^*$. We call this menu an *exclusive-franchise-premium policy*, the premium for exclusivity being $a_M - a_1$.

4.2 Optimal franchising

Our second lemma characterizes the fundamental tension between efficiency and incentive compatibility: the market structure is optimally designed without eliciting the incumbent's private information.

Lemma 3 *: Exclusive-franchise-premium policies (menus) are suboptimal.*

Proof. Consider an exclusive-franchise-premium policy characterized by $(a_1, a_M, a_2(.))$ or alternatively by θ_* and θ^* with $\theta_L \leq \theta_* < \theta^* < \theta_H$ such that the following indifference equations hold for the incumbent:

$$\begin{aligned} \pi_1(\theta_*) - a_1 &= 0 \\ \pi_M(\theta^*) - a_M &= \pi_1(\theta^*) - a_1 \end{aligned}$$

that is, $a_1 = \pi_1(\theta_*)$ and $a_M = \pi_M(\theta^*) + \pi_1(\theta_*) - \pi_1(\theta^*)$. Applying for entry must be individually rational for the entrant, so $a_2(.)$ must satisfy:

$$\int_{\theta_*}^{\theta^*} [\pi_2(\theta) - a_2(\theta)]\, dF(\theta) \geq 0,$$

as well as the budget constraint: for all $\theta \in (\theta_*, \theta^*)$, $a_1 + a_2(\theta) \geq I$ and $a_M \geq I$.

Consider now the policy such that the infrastructure is not built if $\theta < \theta_*$ and it is built and operated under open access if $\theta \in (\theta_*, \theta_H]$. This is obtained by maintaining $a_1 = \pi_1(\theta_*)$ and increasing a_M above $\pi_M(\theta_H)$. Moreover, fix the entrant's access charge so that $a_2 = \min_{\theta \in (\theta_*, \theta^*)} a_2(\theta)$. In the new mechanism, the incumbent ranks decisions 0 and D as before, depending on θ, and has no incentive to ask for an exclusive franchise: so incentive compatibility and individual rationality hold for the incumbent.

Since for θ in (θ^*, θ_H), $W_M(\theta)$ is replaced by $W_D(\theta)$, expected social welfare is unambiguously increased. The budget is balanced since it was for all $\theta \in (\theta_*, \theta^*)$ in the original mechanism. Finally, it is individually rational for the entrant to apply for entry since:

$$\int_{\theta_*}^{\theta^*} [\pi_2(\theta) - a_2] \, dF(\theta) \geq \int_{\theta_*}^{\theta^*} [\pi_2(\theta) - a_2(\theta)] \, dF(\theta) \geq 0.$$

Therefore, the new mechanism satisfies all requirements and yields higher social welfare. ■

The regulator then need not rely on the incumbent's information to select a market structure at the optimum. She can restrict attention to two possible classes of policies.
(1) In the class of *no-competition policies*, the infrastructure is built if and only if $\theta \geq \theta^*$, and in this case it is operated under a monopoly franchise. The highest contribution that can be demanded from the incumbent is: $a_M = \pi_M(\theta^*)$. The optimal no-competition policy is therefore characterized by $\theta^* = \theta_M$, that is the infrastructure is built under monopoly franchise as long as it is financially viable.
(2) In the class of *no-exclusive-franchise policies*, the infrastructure is built if and only if $\theta \geq \theta_*$ and it is then operated under open access. The highest contribution that can be demanded from the incumbent is $\pi_1(\theta_*)$. An argument similar to the one in the proof of the previous lemma shows that the entrant's access charge can be chosen constant and equal to $\mathsf{E}[\pi_2(\theta) \mid \theta \geq \theta_*] = \int_{\theta_*}^{\theta_H} \frac{\pi_2(\theta)}{1-F(\theta_*)} dF(\theta)$. Let the function $a_D(.)$ be defined by:

$$a_D(\theta) = \pi_1(\theta) + \mathsf{E}[\pi_2(z) \mid z \geq \theta]$$

and note that it is increasing in θ. Let θ_D be the unique solution of $a_D(\theta) = I$, i.e. $\theta_D \equiv a_D^{-1}(I)$ or $\theta_D = \theta_L$ if $a_D(\theta) < I$ for all θ. The optimal no-exclusive-franchise policy is characterized by $\theta_* = \theta_D$, with entry fees $a_1 = \pi_1(\theta_D)$ and $a_2 = I - a_1$.

The implementation of this no-exclusive-franchise policy consists in asking the incumbent whether the project should be undertaken or not given his fee $a_1 = \pi_1(\theta_D)$, and in case it should, the entrant is invited to enter for a fee $a_2 = \mathsf{E}[\pi_2(\theta) \mid \theta \geq \theta_D]$. In this procedure, after the incumbent agrees on the project, the entrant should update his beliefs on θ given that $\theta \geq \theta_D$. Given the value of a_2, his decision to enter is then still rational and the policy is robust to this alternative timing of the entrant's decision.

The next proposition summarizes the analysis and states the optimal mechanism:

Proposition 4 *: It is never optimal to give the incumbent a choice between competition with a base contribution and an exclusive franchise for a premium over this base contribution.*

(a) If

$$\int_{\theta_M}^{\theta_H} W_M(\theta)dF(\theta) > \int_{\theta_D}^{\theta_H} W_D(\theta)dF(\theta); \tag{6}$$

the optimal regulatory policy under incomplete information grants the incumbent a monopoly franchise provided the latter is willing to finance the entire investment I. Otherwise, the infrastructure is not built.

(b) If (6) does not hold, the infrastructure is built if and only if the incumbent is willing to contribute $a_1 = \pi_1(\theta_D)$. The entrant then obtains access at $a_2 = \mathsf{E}[\pi_2(\theta) \mid \theta \geq \theta_D]$.

In a nutshell, the regulator cannot screen the incumbent operator to obtain a more competitive structure as demand improves; she must rely on her prior information only. The technically oriented reader will here recognize an instance of non-responsiveness (Guesnerie-Laffont (1984) and Caillaud & al (1988)). Bunching, i.e., the absence of extraction of the agent's information, arises when full information efficiency and incentive constraints require the market structure to move in opposite directions with θ.

Access charges: The definition of θ_D shows that provided open access is optimal, open access is more widespread than under perfect information: $\theta_D < \theta_D^{SI}$, since

$$I = \pi_1(\theta_D^{SI}) + \pi_2(\theta_D^{SI}) < \pi_1(\theta_D^{SI}) + \mathsf{E}\left[\pi_2(\theta) \mid \theta \geq \theta_D^{SI}\right].$$

Assumption A.2 implies that $\pi_M(\theta) \geq \pi_1(\theta)+\pi_2(\theta)$ but it does not imply that $\pi_M(\theta) \geq a_D(\theta)$. So, it is possible that $\theta_D < \theta_M$. In this case, (6) does not hold and the no-exclusive franchise policy is optimal under incomplete information. Then, the infrastructure is built more often than under perfect information and expected welfare is clearly higher than under perfect information. The explanation is obviously that the regulator extracts the entrant's expected profits, which may be higher than the realized profits if market profitability is barely above the threshold level. Entry may then occur only because the entrant is optimistic about θ, while under perfect information the entrant knows θ and is less willing to enter for low θ. As this looks like a pathological situation, we rule it out in the following by assuming:

A.5: Even under incomplete information, competition does not facilitate financing: $a_D(\theta) < \pi_M(\theta)$ for all θ.

Assumption A.5 is for example satisfied when competition in the market is intense, or when the uncertainty about demand is not too large.

Under a no-exclusive-franchise policy, the incumbent and the entrant do not pay the same access charge, even when both firms are technological and commercial equals (so $\pi_1(.) = \pi_2(.) = \pi(.)$), since

$$a_2 = \mathsf{E}\left[\pi(\theta) \mid \theta \geq \theta_D\right] > \frac{I}{2} > \pi(\theta_D) = a_1.$$

There is price discrimination with respect to access. This is no surprise since the incumbent knows θ and earns an informational rent from this knowledge, while the entrant only considers expected profits and gets no expected rent; even though they are technologically similar, both firms are informationally differentiated and should be treated differently.

Remark: It is worth investigating the nature of the optimal access policy when, for exogenous reasons, a non-discrimination constraint were imposed on the infrastructure manager. So, consider a symmetric model where $\pi_1(.) = \pi_2(.) = \pi(.)$ and suppose access must be granted on demand with a nondiscriminatory charge $\hat{a}$. If the access charge is fixed at $\hat{a} = I/2$, then there exists an equilibrium where the incumbent demands access if and only $\theta \geq \hat{\theta}$, where $\hat{\theta}$ is the unique solution of $\pi(\theta) = I/2$ (or equals θ_L), and where entry occurs if and only if the incumbent demands access. This nondiscrimination policy benefits the entrant and hurts the incumbent, but may jeopardize the financing of the infrastructure since $\hat{\theta} > \theta_D$.

4.3 Moderation of competition

Finally, we explore the possibility of moderation of competition that is, of a monopoly-franchise period followed by open access.

Such a possibility is suggested by the World Bank's experience in helping design and structure concession contracts for railways in various countries. For example, for the international link between Abidjan (Côte d'Ivoire) and Ouagadougou (Burkina Faso), a fifteen year concession was set up with a seven year period of exclusivity followed by a period of open access to another operator, after approval by the regulator, for an access fee that was agreed upon at the initial stage (see Campos-Cantos [1999]). Neglecting the (irrelevant) feature that the exclusivity period comes first, this market design can be viewed as a compromise between the two regimes that have been discussed previously, that is a mix between an open access policy and an exclusive franchise policy.

A similar mixed market structure could alternatively be obtained by reserving some rail trackage rights for the exclusive use of an incumbent operator while opening other rights to open access. In Argentina and Brazil, rail concessions have been set up that grant the operator an exclusive right except on certain track segments that are subject to open access.

Technically, our analysis so far has focused on deterministic policies, and we now generalize it to random policies, where the randomness can without loss of generality be interpreted as a temporal switch in the franchising regime.[14]

In a stochastic mechanism, the market structure is determined, conditionally on the infrastructure being built, according to a pre-specified probabilistic decision rule where $x(\tilde{\theta})$ is the probability that a monopoly franchise be granted when the incumbent announces market profitability $\tilde{\theta}$ (and $1 - x(\tilde{\theta})$ the probability that open access obtains).[15] The basic conclusion of Lemma 2 extends to stochastic regulatory policies. Conditionally on building the infrastructure, the incumbent's expected utility as a function of his expected access charge $a_I(\tilde{\theta})$, his announcement $\tilde{\theta}$ and of the true market profitability θ is:

$$U(\tilde{\theta}, \theta) = [\pi_M(\theta) - \pi_1(\theta)]\, x(\tilde{\theta}) + \pi_1(\theta) - a_I(\tilde{\theta}).$$

Standard techniques show that $x(.)$ must be nondecreasing. Thus, the share of access rights reserved for the exclusive use of the incumbent, or the time duration of the exclusive franchise, must increase when market profitability improves. The tension between efficiency and incentive compatibility carries over to stochastic schemes.

As a consequence, there may be a first natural way to try to improve on the optimal deterministic regulatory policy characterized in Proposition 4, namely to propose an open access policy with moderation of competition, corresponding to a mechanism with a constant probability $x \in [0, 1]$ on some interval $[\theta_*, \theta_H]$ where the project is realized. The optimal policy in this class must trade off the social desirability of introducing as much competition as possible with the financial viability constraint that requires to take into account that the sum of the maximal access fee that can be demanded from the incumbent, namely $\{x\pi_M(\theta_*) + (1 - x)\pi_1(\theta_*)\}$, and of the maximal entry fee that can be imposed on the entrant, namely $(1 - x)\mathsf{E}\,[\pi_2(z) \mid z \geq \theta_*]$, must cover the investment cost;

[14] If T is the length of the monopoly franchise and r the rate of interest, then $x = \left(1 - e^{-rT}\right)$ and $1 - x = e^{-rT}$ are the probabilities in the equivalent stochastic mechanism.

[15] Actually, a fully stochastic mechanism would also imply some randomization about the decision to build the infrastructure or not. We do not consider this possibility.

that is, it solves the following program:

$$\max_{\{x,\theta_*\}} \int_{\theta_*}^{\theta_H} [xW_M(\theta) + (1-x)W_D(\theta)]\, dF(\theta)$$
$$\text{s.t.} \quad : \qquad x\pi_M(\theta_*) + (1-x)a_D(\theta_*) \geq I.$$

While Proposition 4 compares the polar cases $x = 0$ and $x = 1$, the Appendix shows that the solution of this program may be interior, with $x_* \in (0,1)$. Proposition 5 summarizes the analysis (see the Appendix for further developments). Part (i) refers to arbitrary stochastic mechanisms, while part (ii) uses constant probability mechanisms.

Proposition 5 *: (i) Any stochastic incentive compatible mechanism is characterized by a threshold level θ_* and a nondecreasing probability of granting exclusivity rights $x(.)$ such that:*

- *if $\theta \in [\theta_L, \theta_*)$, the infrastructure is not built;*
- *if $\theta \in (\theta_*, \theta_H]$, the infrastructure is built and operated under a monopoly franchise with probability $x(\theta)$ and under open access with probability $1 - x(\theta)$.*

(ii) In certain environments, it is possible to improve upon an indefinite monopoly franchise or an indefinite open access policy through the use of moderation of competition.

The general message is clear: eliciting information from the incumbent is of little help in designing the optimal market structure and the optimal degree of competition. We return to deterministic mechanisms in the rest of the paper. The insights derived in this section would apply to the next sections as well.

5 Overall budget constraint and cross-subsidies among investment projects

We have assumed that the regulator is instructed to fully fund the investment cost *of any given project*. This section shows that the result that incomplete information drastically impacts the downstream market design is robust to different formalizations of the regulatory mandate. Ex post budget balance, defined on a project by project basis, is a strong and restrictive rule. While in the railroad industry this mandate applies to RFF in France, and to a certain extent in Germany, we must consider alternative regulatory mandates and their impact on downstream market design under incomplete information.

This section assumes that the regulator must cover the costs of its investments *on average* over its various investment projects. Thus the regulator is allowed to subsidize some projects as long as the shortfall is covered by surpluses on other projects. There is a large number (technically a continuum of mass one) of projects, indexed by $\theta \in [\theta_L, \theta_H]$, where θ characterizes the value of an individual project. Each individual project involves investment cost I. The regulator cannot tell the projects apart, and simply knows the distribution of projects, summarized by the c.d.f. $F(.)$. Each project is as described earlier, with an incumbent who knows the parameter θ, and a potential entrant. As earlier, the regulator can select information contingent decision rule $\delta(\theta) \in \{O, M, D\}$ and access charges.

5.1 Same incumbent on all segments: the rotten segments curse and optimal cross-subsidies

When the incumbent operator is the same in all markets, by the law of large number, the incumbent cannot fool the regulator over the aggregate distribution of the demand parameters.[16] The incumbent therefore enjoys no rent. The regulator fixes the relative proportions of the various market structures. Formally, it may specify the fraction of projects m_0 that will be rejected, the fraction of projects m_D that will be implemented under open access and the fraction of projects m_M that will be implemented under a monopoly franchise, with $m_0 + m_D + m_M = 1$.

While the incumbent has no private information about the distribution of projects, he has superior information about each particular project. Given a *total* access charge A over all segments, he selects which investments are made under a monopoly franchise,

[16] This would not be so if the incumbent had private information about the distribution, and not only about the value of individual projects.

which are made under open access, and which are let down:

$$\max_{\{\Theta_D \Theta_M\}} \left[\int_{\Theta_D} \pi_1(\theta) dF(\theta) + \int_{\Theta_M} \pi_M(\theta) dF(\theta) - A \right] \tag{7}$$

s.t.

$$\int_{\Theta_D} dF(\theta) = m_D \quad \text{and} \quad \int_{\Theta_M} dF(\theta) = m_M \ ,$$

where Θ_D is the set of types of projects that the incumbent chooses to have implemented under open access, and Θ_M the set over which he chooses to enjoy operating monopoly. The incumbent chooses the monopoly regime for the most profitable projects, in the limit of a fraction m_M of projects, and then the next profitable projects under open access, in the limit of a fraction m_D of projects; he rejects the remaining, lowest profitability projects. More formally, $\Theta_M = (\theta^+, \theta_H)$, $m_M = 1 - F(\theta^+)$ and $\Theta_D = (\theta_-, \theta^+)$, and $m_D = F(\theta^+) - F(\theta_-)$. The choice of (m_D, m_M) is then equivalent for the regulator to the choice of two threshold levels as in previous sections and incentive compatibility has the same implications as earlier: high profitability projects are built under a monopoly franchise while intermediate profitability projects are built under open access. This we call the "rotten segments curse", since the regulator is very constrained in creating competition on the most profitable segments.

The incumbent's access charge can be determined only on a global basis, and leaves the incumbent with no rent:

$$A = \int_{\theta_-}^{\theta^+} \pi_1(\theta) dF(\theta) + \int_{\theta^+}^{\theta_H} \pi_M(\theta) dF(\theta),$$

and so the average budget constraint over all projects takes the simple form:

$$\int_{\theta_-}^{\theta^+} (\pi_1(\theta) + \pi_2(\theta) - I)\, dF(\theta) + \int_{\theta^+}^{\theta_H} (\pi_M(\theta) - I) dF(\theta) \geq 0. \tag{8}$$

The regulator maximizes (average) social welfare,

$$\int_{\theta_-}^{\theta^+} W_D(\theta)dF(\theta) + \int_{\theta^+}^{\theta_H} W_M(\theta)dF(\theta)$$

subject to (8).

Note that this program is formally equivalent to the program that would obtain if the regulator did not impose an average budget constraint but could implement transfers from taxpayers to the regulator with a cost of public funds λ. The only difference is that λ would then be a given characteristics of the whole economy, while it is here the endogenous Lagrange multiplier associated with the budget constraint.

Letting λ denote the shadow price of the budget constraint and assuming interior solutions, the first-order necessary conditions are:

$$W_D(\theta_-) = \lambda[I - \pi_1(\theta_-) - \pi_2(\theta_-)], \quad (9)$$

and

$$W_D(\theta^+) - W_M(\theta^+) = \lambda[\pi_M(\theta^+) - \pi_1(\theta^+) - \pi_2(\theta^+)]. \quad (10)$$

The interpretation of these conditions is straightforward. Accepting more projects (reducing θ_-) creates a deficit with shadow cost $\lambda[I - \pi_1(\theta_-) - \pi_2(\theta_-)]$ but generates welfare $W_D(\theta_-)$. And similarly, an expansion of the monopoly region (a reduction of θ^+) generates a budget surplus with shadow value $\lambda[\pi_M(\theta^+) - -\pi_1(\theta^+) - \pi_2(\theta^+)]$, but reduces welfare by $W_D(\theta^+) - W_M(\theta^+)$.

Proposition 6 *: Suppose that the regulator must break even only over its entire network, that the incumbent is the same on all segments, and that the incumbent has information about market demand on each segment. Then the most profitable segments are served under a monopoly franchise. The segments under a monopoly franchise cross-subsidize those operated under open access. And some open access investments are made that do not break even on an individual basis.*

5.2 Multiple incumbents

Let us now entertain the opposite hypothesis that each segment is initially served by a segment-specific incumbent (so, technically, there is a "continuum of incumbents"). The essential difference with the single-incumbent case is that each incumbent's utility is unknown, so that the incumbents enjoy rents. The analysis on the incumbents' side is identical to that of sections 3 and 4. The new feature relative to these sections is again that the regulator can lose money on some investments and recoup it on others.

Would the regulator want to treat otherwise identical segments differently? Let m_D (respectively, m_M) denote the fraction of segments for which the regulator offers open access (respectively, a monopoly franchise), where, without loss of generality,

$$m_D + m_M = 1.$$

Let $\hat{\theta}_D$ and $\hat{\theta}_M$ denote the lowest types for which the investment is made for a segment designated as an open access (respectively, a monopoly franchise) segment. The access charges on the two types of segments are therefore $\pi_1(\hat{\theta}_D)$ and $\pi_M(\hat{\theta}_M)$.

The regulator's program is

$$\max_{\{\hat{\theta}_M, \hat{\theta}_D\}} \left\{ m_M \int_{\hat{\theta}_M}^{\theta_H} W_M(\theta)dF(\theta) + m_D \int_{\hat{\theta}_D}^{\theta_H} W_D(\theta)dF(\theta) \right\}$$

s.t.

$$m_M \left[1 - F(\hat{\theta}_M)\right]\left[\pi_M(\hat{\theta}_M) - I\right] + m_D \left[1 - F(\hat{\theta}_D)\right]\left[a_D(\hat{\theta}_D) - I\right] \geq 0. \qquad (11)$$

Let us now make:[17]

A.6 $\frac{\dot{\pi}_M(\hat{\theta}_M)}{W_M(\hat{\theta}_M)} \geq \frac{\dot{a}_D(\hat{\theta}_D)}{W_D(\hat{\theta}_D)}$ *whenever* $\hat{\theta}_D \geq \hat{\theta}_M$.

[17] Assumption A.6 is essentially an assumption about the shape of the distribution function $F(.)$ and the rate of increase of π_2. It is easy to see that:

$$\dot{a}_D(\theta) = \dot{\pi}_1(\theta) + \frac{f(\theta)}{1 - F(\theta)} \cdot \int_{\theta}^{\theta_H} \frac{[1 - F(s)]}{[1 - F(\theta)]} \dot{\pi}_2(s)ds.$$

So, A.6 is satisfied if both π_2 and the hazard rate are relatively insensitive to variations in θ.

Assumption A.6 is a relatively mild assumption. Noting that $W_M(\hat{\theta}_M) \leq W_D(\hat{\theta}_D)$, then $\dot{\pi}_M(\hat{\theta}_M) \geq \dot{a}_D(\hat{\theta}_D)$ is sufficient for A.6 to hold. Since a_D is closely related to total duopoly profit, this sufficient condition *roughly* means that increases in demand lead to higher increases in profit under monopoly. It is satisfied for example in the standard Hotelling model,[18] or when uncertainty about θ is small and monopoly profit increases faster with demand than duopoly profit (which it does in standard models). The following proposition is demonstrated in the Appendix:

Proposition 7 *: Suppose that the budget constraint is over the entire network, that segments are served by different incumbents who know the demand on their own market, that A.6 holds, and that, in the single-segment context, the two institutions (open access, monopoly franchise) deliver roughly equivalent welfares. Then,*
(i) it is optimal to discriminate among otherwise identical segments and to operate some under open access and others under a monopoly franchise;
(ii) the monopoly segments cross-subsidize open access ones ($\pi_M(\hat{\theta}_M) - I > 0 > a_D(\hat{\theta}_D) - I$).

6 Dynamics

Let us now extend the basic model in yet another direction, that of repeated licensing. We consider the single-project context in a two-period setup ($t = 1, 2$) and assume that the relationship between regulator and the operator(s) is run by short-term licenses; we will analyze the potential gains from long-term contracting. We also assume that the project must break even each period.

Two issues are worth of study. First, the incumbent's realization that his current acceptance impacts the regulator's beliefs about demand[19] and therefore the latter's

[18] Let θ denote the uniform mass of consumers along the interval $[0, 1]$ and let there be two products at the two extremes of the segment. Assuming the market is covered and letting v, t and c denote consumer valuation, transportation cost and marginal cost, $\pi_M(\theta) = \theta \, v - \frac{t}{2} - c$ and $a_D(\theta) = \frac{\theta t}{2} + \frac{\theta + \theta_H}{2} \frac{t}{2}$. Since $v - \frac{t}{2} - c > t > \frac{3t}{4}$, the assumption is satisfied.

[19] We maintain our assumption that the regulator does not observe profits. Were the regulator to regulate the incumbent's profit on this segment (which, recall, may not be an easy task due to potential

future policy may lead him to modify his strategy. This ratchet effect has been analyzed extensively,[20] but its implications for the particular context at hand haven't been derived. Second, current entry may allow future "benchmarking" by allowing another operator to acquire information about market demand.

6.1 Ratcheting

To separate the two issues, let us first assume that if there is entry at date 1, the entrant doesn't learn demand before date 2.[21] So, even in case of open access at date 1, the incumbent still has an informational advantage at the beginning of date 2, and so entry does not bring about any benchmarking benefit.

In this setup, it is well-known[22] that the optimal long-term contract is the repetition of the optimal static contract, here described in proposition 4. This implies that if the incumbent finds it optimal not to strategically alter its behavior when facing a short-term contract of open access or monopoly franchise, then the lack of commitment imposes no cost. Let us recall that under assumption A.5, $\theta_M \leq \theta_D$ and consider the two cases envisioned in proposition 4:

(a) *Monopoly franchise is optimal in the static context.*

Suppose the regulator offers the optimal static contract at date 1, that is, proposes a short-term monopoly franchise at fee $a_M = \pi_M(\theta_M) = I$ to the incumbent. We claim that the dynamic perspective does not alter the incumbent's behavior, and that the regulator offers at date 2 to renew the monopoly franchise at the same fee a_M.[23] From a myopic viewpoint, the incumbent should accept the franchise if and only if $\theta \geq \theta_M$. Then,

cross-subsidies with the incumbent's other activities), the incumbent would further have an incentive for underprovision of effort at date 1 or for delaying income recognition (either through accounting methods or by investing in customer lock in) in order to signal a low demand on this particular segment.

[20] See, e.g., Freixas et al (1985) and Laffont-Tirole (1988).

[21] This obviously is a very strong assumption. On the other hand, it is reasonable to assume that the entrant's informational handicap does not vanish instantaneously after entry. This can be justified either by the presence of noise or by delays in income recognition.

[22] See, e.g., Baron-Besanko (1984) and Caillaud et al (1988).

[23] Provided the incumbent has accepted the monopoly franchise at date 1. Otherwise, the regulator learns that $\theta < \theta_M$ and therefore offers no contract at date 2 (or could, alternatively and without modification, offer the monopoly franchise at fee a_M).

the regulator's posterior beliefs when the incumbent has accepted the franchise is the truncated distribution $f(\theta)/[1-F(\theta_M)]$ on $[\theta_M, \theta_H]$. The regulator therefore offers the same monopoly franchise at date 2. And since the second-period fee must be at least equal to I, the incumbent will never be able to get better conditions in period 2.

We therefore conclude that monopoly franchising is immune to the ratchet effect. We now show that the same is not true for open access.

(b) *Open access is optimal in the static context.*

Suppose that at date 1 the regulator offers open access at access charge $a_1 = \pi_1(\theta_D)$ to the incumbent, and that, if the incumbent accepts, the regulator offers the same open-access-at-access-charge a_1 license at date 2. The hypothesized property that the open access policy is immune to the ratchet effect means that the incumbent accepts the duopoly franchise at date 1 if and only if $\theta = \theta_D$. Note in particular that type $\theta = \theta_D$ obtains no rent in either period.

Suppose now that the incumbent refuses the franchise at date 1. Then the regulator learns that $\theta < \theta_D$, and therefore offers the monopoly franchise at fee $a_M = \pi_M(\theta_M)$ at date 2, resulting in particular in a rent $\delta[\pi_M(\theta_D) - \pi_M(\theta_M)] > 0$ for type θ_D, where δ is the discount factor.

We therefore conclude that the incumbent may have an incentive to convince the regulator that the demand forecast was too optimistic and that this market is a natural monopoly. The open access policy is not immune to the ratchet effect.

To analyze equilibrium behavior when open access is offered at date 1, let us assume that the discount factor is not large:

A.7 : $\quad \dot{\pi}_1(\theta) > \delta\dot{\pi}_M(\theta) \quad \text{for all} \quad \theta.$

Assumption A.7 means that discounting is large enough that as demand grows an open access license today matters more and more to the incumbent relative to a monopoly franchise tomorrow.

Suppose that at date 1 the regulator offers open access at access charges a_1 and a_2 for the incumbent and the entrant. Assumption A.7 implies that in equilibrium, the incumbent accepts the open access license if and only if $\theta \geq \hat{\theta}$ for some $\hat{\theta}$ to be determined. The open access financing condition therefore becomes:

$$a_1 + \mathsf{E}\left[\pi_2(\theta) \mid \theta \geq \hat{\theta}\right] \geq I. \tag{12}$$

Because accepting the license signals high realizations of demand and therefore can only reduce the incumbent's date-2 welfare, necessarily

$$a_1 \leq \pi_1(\hat{\theta}),$$

and so

$$\hat{\theta} \geq \theta_D.$$

To simplify the analysis without loss of insights, let us assume that (6) is satisfied with equality, so a monopoly franchise and open access are equivalent in the single-project static context. Then a rejection of the open access license induces a monopoly franchise offer (at fee a_M) at date 2, while an acceptance leads to a new open source license at fee $\pi_1(\hat{\theta})$ for the incumbent, implying that type $\hat{\theta}$ does not obtain any second-period rent.[24] Therefore

$$\pi_1(\hat{\theta}) - a_1 = \delta\left[\pi_M(\hat{\theta}) - \pi_M(\theta_M)\right]. \tag{13}$$

The break-even condition under open access therefore becomes more stringent as δ grows:

$$a_1 + \mathsf{E}\left[\pi_2(\theta) \mid \pi_1(\theta) - \delta\left[\pi_M(\theta) - \pi_M(\theta_M)\right] \geq a_1\right] = I. \tag{14}$$

[24] After acceptance by the incumbent, the regulator has in fact some discretion in the choice of access fees for the incumbent and the entrant, since $\hat{\theta} > \theta_D$ implies that $a_D(\hat{\theta}) > I$. We assume the regulator chooses the largest access charge that is acceptable by the incumbent in period 2.

As the discount factor grows, the open-access cutoff $\hat{\theta}$ increases. This reduces welfare at date 1 (lack of production on $[\theta_D, \hat{\theta}))$ and at date 2 (lack of competition on the same interval).

We can summarize our analysis in

Proposition 8 *: (i) The monopoly franchise is immune to the ratchet effect (ii) In contrast, an absence of long-term commitment makes the open access policy less attractive, as the incumbent attempts to signal a low profitability of the segment. Under assumption A.7, the incumbent accepts the open access license if and only if $\theta \geq \hat{\theta}$, where $\hat{\theta} > \theta_D$ increases with the discount factor.*

6.2 Future benchmarking benefits of open access

Section 6.1 showed that, in the absence of benchmarking benefit, repeated interaction strengthens the case for a monopoly franchise. We investigate whether this conclusion is affected by the introduction of benchmarking benefits. Let us therefore entertain the polar assumption that in case of date 1 competition the entrant learns the state of demand at the end of date 1. The regulator can then fully extract rents at date 2.[25] Maintaining assumption A.7, we see that the incumbent accepts the open access license if and only if $\theta \geq \hat{\theta}$, where $\hat{\theta}$ is still given by (13). The analysis of section 6.1 therefore remains valid.

Proposition 9 *: In the single-project context, proposition 8 still holds when the entrant learns the state of demand perfectly at date 1: There is no benchmarking benefit of open access.*

Proposition 9 hinges on our assumption that the regulator attaches no intrinsic value to extracting the incumbent's rent. It does not carry over to the case of multiple projects, multiple incumbents and a network-wide budget constraint (see section 5.2). Then, extracting the rent created by favorable realizations of demand generates a budget surplus

[25] This results from Maskin (1999). The benefits of benchmarking in a regulatory context have been studied by Caillaud (1990) and Shleifer (1985).

which can be used fruitfully to cross-subsidize other projects. Open access may then become more, rather than less, attractive in a dynamic context.

7 Capture and the open access presumption

We have assumed non-cooperative behavior from all actors, in particular that the regulator does not collude with the operators. Let us now relax this assumption. As argued in Laffont-Tirole (1991), regulatory capture is associated with private information and the existence of potential rents. This implies here that regulatory capture, if it occurs, is capture by the incumbent. The interesting question is whether the concern about regulatory capture favors open access or monopoly franchising.

Suppose the regulator is only concerned about its self-financing mandate and that there is some uncertainty about the benefits of competition. The regulator may have some information about these benefits, on which it would be socially valuable to base the market design decision. The incumbent may however prefer this information not to be disclosed if it is detrimental to him, hence the possibility of capture and the necessity to curb it. More specifically, suppose there are two states of nature: in state $\omega = D$, which occurs with probability α, welfare under competition is higher than in the other state $\omega = d$: $W_D(\theta) \geq W_d(\theta)$ for all θ.

The state of nature could be due to uncertainty about the intensity of competition in duopoly (how likely it is that the firms will engage in tacit collusion) or about the entrant's entry cost (which could be the same as the incumbent's, except that the latter is already sunk). Let us follow this second interpretation so that the state of nature does not affect the incumbent's profits (under either regime); for the entrant, $\pi_{2D}(\theta) > \pi_{2d}(\theta)$ for all θ, the difference being equal to the difference in sunk costs. Following Proposition 4, let θ_D and θ_E be the optimal cut-off points when there is evidence that $\omega = D$ and when there is no evidence on ω:

$$\pi_1(\theta_D) + \mathsf{E}\left[\pi_{2D}(y) \mid y \geq \theta_D\right] = I$$

$$\pi_1(\theta_E) + \mathsf{E}\left[\alpha\pi_{2D}(y) + (1-\alpha)\pi_{2d}(y) \mid y \geq \theta_E\right] = I.$$

It follows that $\theta_D < \theta_E$ and correspondingly, the incumbent's access charge under competition satisfies:

$$a_{1D} \equiv \pi_1(\theta_D) < a_{1E} \equiv \pi_1(\theta_E).$$

Finally, let us also assume that Assumption A5 still holds so that $\theta_M < \theta_D$.

With probability ξ, the regulator has perfect information about the increase in consumer surplus brought about by competition. When the regulator has such information, this information is hard (may be disclosed), and it is shared by the incumbent. The entrant, however, is never informed about the state of nature. To make things interesting, let us assume that open access is optimal (when the signal is common knowledge, that is in the sense of proposition 4) in state D, while a monopoly franchise is optimal in state d.

Suppose, first, that in the absence of signal and when there is no threat of collusion, a duopoly franchise is optimal. Then capture is not an issue: Open access will prevail unless the regulator offers evidence that $\omega = d$, in which case a monopoly franchise is granted to the incumbent.[26] Because the default option in the absence of disclosure is open access, the incumbent has no incentive to capture the regulator.

The interesting case is then when absent any signal society would prefer a monopoly franchise. Intuitively, the incumbent is then eager to capture the regulator in case of a pro-competitive signal (D), since the regulator is indifferent between disclosing or hiding the signal D provided the project is fully funded, while the incumbent's informational rent is larger when the signal D is not disclosed. The incumbent's gain from hiding the signal D is given by:

$$\Delta(\theta) = [\pi_M(\theta) - \pi_M(\theta_M)] - \sup\{\pi_1(\theta) - \pi_1(\theta_D); 0\} > 0.$$

[26] More precisely, the default option is open access with access charges a_{1E} and $a_{2E} = I - a_{1E}$. If evidence is provided that $\omega = D$, then open access still prevails, but with access charges a_{1D} and $a_{2D} = I - a_{1D}$.

Thus, the policy that is optimal under a public signal always induces monopoly franchising when the benefits of competition is private knowledge.

To prevent capture, two options are open.[27] First, and least realistically, the regulator may be rewarded for creating open access by making a convincing case for it (revealing D). The policy that can be implemented then coincides with the optimal policy with public signal, but welfare is reduced by an additional cost of $\xi\alpha\Delta(\theta_H)$ (in expected terms), that corresponds to the regulator's reward.[28] The regime is then not affected by the threat of capture.[29]

Second, one may eliminate the stake in collusion by giving the benefit of the doubt to competition. That is, the regulator is required to offer open access unless he makes a convincing case in favor of monopoly.[30] (or in favor of open access with access charges a_{iD}). In technical terms, open access prevails in the absence of signal (even though a monopoly franchise is then optimal in the absence of potential capture), and a monopoly franchise is granted only when the regulator reveals evidence that $\omega = d$. There is no scope for collusion anymore, but expected welfare is reduced, compared to the case of public signal, by the fact that monopoly franchise is chosen when the regulator does not get information, while open access would be preferable.

If ξ approaches 1, however, this social cost vanishes since it is unlikely that the regulator does not get information about ω. On the other hand, capture involves a non-negligeable social cost since a monopoly franchise is decided in all occurrences where the regulator has evidence that open access would be optimal (with probability $\xi\alpha$). Simi-

[27] In this intuitive argument, we restrict attention to capture-proof policies that is, to policies that prevent collusion for sure. But the insight holds more generally.

[28] One can formulate alternative assumptions as to whether this cost enters the budget constraint. Because this case is not the focus of the section, we will not expand on this further.

[29] Note that preventing collusion with probability one, that is for all realizations of θ, might be excessively costly, depending upon the form of the welfare and profit functions. It might be socially preferable to adopt a policy that allows collusion to develop when it is too costly to fight. For instance, fixing a level of reward s for the infrastructure owner such that $\Delta(\hat{\theta}) = s$ will deter collusion except for $\theta > \hat{\theta}$, in which case an exclusive franchise will be recommended by the infrastructure owner while open access would be preferable.

[30] Again, the default option is open access with access charges a_{iE}, and if evidence is provided that $\omega = D$, access charges are a_{iD}.

larly, the social cost of rewarding the regulator is non-negligeable (proportional to $\xi\alpha$). We summarize this discussion in the following proposition.

Proposition 10 *: When the possibility of capture is a concern and the regulator is likely to be informed about consumer benefits (that is when ξ is large enough), the optimal capture-proof policy is to impose open access whenever no clear evidence is provided in favor of exclusivity.*

Note that, as is often the case when the institutional response to the threat of capture is accounted for, the incumbent is actually hurt by the suspicion that he might attempt to capture the regulator. The general message is that the concern about regulatory capture can only improve the case in favor of no-exclusive-franchise policies.

This conclusion does not depend upon the specific assumption that uncertainty concerns the entrant's fixed cost of entry. Taking the alternative interpretation in terms of intensity of competition would lead us to assume that $\pi_{iD}(\theta) < \pi_{id}(\theta)$ for $i = 1, 2$ and all θ. Assuming that A.5 holds for all ω, the incumbent would still try to prevent disclosure of signal D to benefit from a monopoly franchise. The additional difficulty comes from the fact that it is possible that the incumbent's access fee under the no-exclusive-franchise policy is larger when there is evidence about D than when no evidence is available. In such a case, the incumbent prefers to prevent disclosure of signal D even if the default option is open access, as this enables him to pay a smaller access charge under a duopoly regime. The conclusion of the previous proposition is still valid, namely that market design should be biased toward open access to fight capture, but it may be that capture to manipulate the access charges has to be deterred through additional instruments.

8 Further extensions

8.1 Transfer of the monopoly franchise

So far, we have ruled out the possibility that the regulator removes the incumbent and grants a monopoly franchise to the entrant. This section extends the model to deal with the possibility of franchise transfer.

To analyze the possibility of replacing the incumbent, consider the setting of Sections 3 and 4, with ex post budget balance. Let $\pi_E(\theta)$ denote the entrants profit as a newly franchised monopoly, and $W_E(\theta)$ the corresponding social welfare. To fix ideas, suppose that the incumbent has some special know-how or experience, or that the transfer of the franchise involves transaction/moral hazard costs, or else that consumer brand recognition is lost in the process of transferring the license, so that first-best efficiency as well as monopoly profits are higher when the incumbent, rather than the entrant, is the franchised monopoly. Formally, for all θ,

$$\pi_1(\theta) + \pi_2(\theta) < \pi_E(\theta) < \pi_M(\theta)$$

$$W_E(\theta) < W_M(\theta) < W_D(\theta).$$

We naturally assume that: $\dot{\pi}_E(\theta) > 0$. The model of sections 3 and 4 can be viewed as a special case of this more general model, in which $W_E(\theta)$ is negative.

The perfect information optimal policy is the same as in Section 3. Open access is optimal provided it allows self financing ($\theta > \theta_D^{SI}$), that is for high value projects. When $\theta < \theta_D^{SI}$, it is socially preferable to let the incumbent, rather than the entrant, be the monopolist because allocating the franchise to the incumbent both yields a higher social welfare and facilitates financing. So, there is no franchise transfer and the incumbent is never replaced.

Under incomplete information, the incentive analysis is very similar to the one performed in section 4, except that we can now introduce an access fee a_E for the entrant under monopoly and some compensation payment $T \geq 0$ for the incumbent in case of franchise transfer. The incumbent monopolist must now take into account the possibility of transferring the franchise to the entrant, which yields him T. Assuming that the infrastructure is not built for low realizations of demand, necessarily $T = 0$.

The analysis in Section 4 is not modified if the regulator chooses the monopoly franchise, since for all $\theta < \theta_M$, $\pi_E(\theta) < \pi_M(\theta) \leq I$. More generally, however, if the regulator does not give a monopoly franchise to the incumbent, the possibility of granting a franchise to the entrant should be taken into account on some range (θ_E, θ_*) with $a_E = E[\pi_E(\theta) \mid \theta \in (\theta_E, \theta_*)]$ provided this is larger than I. Following the analysis in Section 4, the infrastructure is built and operated under open access for $\theta \in [\theta_*, \theta_H]$. Then, if $\pi_E(\theta_D) > I$, there exists a unique threshold θ_E, such that $\theta_L \leq \theta_E < \theta_D$ and

$$\mathsf{E}[\pi_E(\theta) \mid \theta_E \leq \theta \leq \theta_D] = I. \tag{15}$$

The range of operations can then be extended to $[\theta_E, \theta_H] \supseteq [\theta_D, \theta_H]$ while meeting the budget constraint thanks to the entrant's monopoly profits. But it may be optimal to let $\theta_* > \theta_D$. While this reduces welfare for $\theta \in (\theta_*, \theta_D)$, this policy may allow an extension of operations in the lower end of the distribution of the demand parameter.

The franchise transfer policy is straightforward to understand. It has one important consequence, in terms of policy. It is now possible that an intermediate value project be realized under monopoly franchise, as under perfect information; but then, the franchise should be granted to the entrant, not to the incumbent. The franchise transfer policy has also one important weakness: the regulator must trust the incumbent to cooperate when he exits the market, even though he has no strict incentive to reveal that the infrastructure should be operated under an entrant franchise versus not be built at all. It would be worth developing a model where the incumbent is not indifferent between these two alternatives (transfer the monopoly franchise or let go the project).

8.2 Profit observability

Our assumption that the profit on the competitive segment is not regulated is fine for some concessions and regulated industries, but not for others. The regulator's ability to observe profit calls for earnings sharing schemes, in which some of the incumbent's profit is passed through to consumers. The extent of passthrough depends on the extent of moral hazard, broadly construed to include both X-inefficiency and cost-padding or more generally cost manipulations. For example, only limited amounts of passthrough are feasible and our analysis extends without modifications if the incumbent can easily transfer income or resources between the regulated segment and the unregulated activities (or segments regulated by other authorities, as in the case of water concessions awarded at the local level).

More generally, the key robustness issue is whether, under profit sharing, the incumbent still finds it relatively more profitable to obtain an exclusive franchise for high demand (or low cost). The "rotten segment curse" might no longer obtain if incentives were designed so as to be high-powered (involve low passthrough) in the open access region and low-powered (involve high passthrough) in the exclusive franchise region. Then, an increase in demand would be more profitable under open access than under an exclusive franchise, reversing assumption A.4 and substantially complicating the analysis. Again, it is feasible to find sufficient conditions for this reversal not to happen and therefore for the analysis of this paper to carry over to profit observability. For example, it suffices that moral hazard be socially very costly (profit sharing leads to a sharp increase in cost); or that uncertainty about θ be low enough; or else, that product market competition under open access be very intensive (close to Bertrand competition), so that $\pi_1(\theta)$ grows slowly with θ and so, offsetting the basic effect embodied in assumption A.4 requires completely destroying incentives under exclusive franchising.

9 Conclusion

Infrastructure projects are often confronted with large uncertainty about market demand and possibly operating cost. Authorities in principle can attempt at eliciting incumbent firms' information on this aspect by offering a basic open access package combined with an exclusivity option in exchange of a higher investment contribution. Alas, the incumbent firm will demand exclusivity precisely when competition is socially most desirable. Authorities therefore cannot screen incumbents to select among open access, a monopoly franchise, or a moderation of competition.

Building on this basic insight, we showed that with multiple projects, segments operated under a monopoly franchise must cross-subsidize open access segments. We then showed that monopoly franchising is immune to the ratchet effect while open access is not; and that a concern for capture makes open access a more appealing default rule.

Needless to say, this first investigation of endogenous market structure with common values leaves scope for future theoretical inquiry. Furthermore, the implications of the present inquiry in different contexts are definitely worth drawing. As we noted in the introduction, competition authorities in their application of the essential facility doctrine confront a similar financing-vs-competition trade-off and lack of information about market viability as regulators of telecommunications, energy or transportation services. The influence of "ex post intervention" rather than "ex ante regulation" is worth thinking about. Similarly, our analysis could be useful when thinking about more sophisticated patent systems in which innovators would pay a premium for increased protection from competition (increased patent length, patent breadth, or freedom in licensing). These topics and others are left for future research.

References

[1] Armstrong, M., Cowan, S., and J. Vickers (1994) *Regulatory Reform: Economic Analysis and British Experience*, Cambridge: MIT Press.

[2] Auriol, E., and J.J. Laffont (1992) "Regulation by Duopoly," *Journal of Economics and Management Strategy*, 1: 507–533.

[3] Baron, D., and D. Besanko (1984) "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1:447–470.

[4] Caillaud, B. (1990) "Regulation, Competition and Asymmetric Information," *Journal of Economic Theory*, 52: 87–110.

[5] Caillaud, B., Guesnerie, R., Rey, P., and J. Tirole (1988) "Government Intervention in Production: A Review of Recent Contributions," *Rand Journal of Economics*, 19: 1–26.

[6] Campos, J. and P.Cantos, (1999) "Regulating Privatized Rail Transport", Worldbank Policy Research WP 2064.

[7] Dana, J., and K. Spier (1994) "Designing a Private Industry: Government Auctions with Endogenous Market Structure," *Journal of Public Economics*, 53(1): 127–47.

[8] Engel, E., Fischer, R., and A. Galetovic (1997) "Highway Franchising: Pitfalls and Opportunities," *American Economic Review, Papers and Proceedings*, 87: 68–72.

[9] —— (1998) "Least-Present-Value-of-Revenue Auctions and Highway Franchising," mimeo, University of Chile.

[10] Estache, A. (1999), "Privatization and Regulation of Transport Infrastructure in the 1990s", World Bank Policy Research WP 2248.

[11] Freixas, X., Guesnerie, R., and J. Tirole (1985) "Planning under Incomplete Information and the Ratchet Effect," *Review of Economic Studies*, 52: 173–192.

[12] Guesnerie, R., and J.J. Laffont (1984) "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25: 329–369.

[13] Kerf, M., D.Gray, T.Irwin, C.Levesque and R.Taylor (1998) "Concessions for Infrastructure", *World Bank Technical Paper N°*399.

[14] Laffont, J.J., and J.Tirole (1988) "The Dynamics of Incentive Contracts," *Econometrica*, 56: 1153–1175.

[15] —— (1991) "The Politics of Government Decision-Making: A Theory of Regulatory Capture," *Quarterly Journal of Economics*, 106: 1089–127.

[16] —— (1993) *A Theory of Incentives in Procurement and Regulation,* Cambridge: MIT Press.

[17] —— (1999) *Competition in Telecommunications*, Cambridge: MIT Press.

[18] Maskin, E. (1999) "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66(1): 23–38.

[19] Milgrom, P. (1996) "Procuring Universal Service: Putting Auction Theory to Work," lecture at the Royal Swedish Academy of Sciences, Dec. 9.

[20] Shleifer A. (1985) "A Theory of Yardstick Competition," *Rand Journal of Economics*, 16: 319–327.

A Appendix 1: Stochastic mechanisms

This appendix offers a technical discussion of the stochastic mechanisms considered in section 4.3.

An incentive compatible mechanism is characterized by $(x(.), \theta_*)$ such that $x(.)$ is non-decreasing. The full description of a mechanism involves access charges: $a_M(.)$ and $a_1(.)$ for the incumbent under respectively exclusivity and open access and $a_2(.)$ for the entrant (under open access, obviously). The integral form of the incentive constraint implies that for all $\theta \geq \theta_*$,

$$x(\theta)a_M(\theta) + (1 - x(\theta))a_1(\theta) = x(\theta)\pi_M(\theta) + (1 - x(\theta))\pi_1(\theta) - \int_{\theta_*}^{\theta} x(t)\left[\dot{\pi}_M(t) - \dot{\pi}_1(t)\right] dt.$$

Budget balance requires:

$$x(\theta)a_M(\theta) + (1 - x(\theta))\left[a_2(\theta) + a_1(\theta)\right] \geq I.$$

Finally, ex ante acceptability for the entrant requires:

$$\int_{\theta_*}^{\theta_H} (1 - x(\theta))\left[\pi_2(\theta) - a_2(\theta)\right] dF(\theta) \geq 0.$$

It is clear that only $a_I(\theta) \equiv x(\theta)a_M(\theta) + (1 - x(\theta))a_1(\theta)$ matters. So,

$$\int_{\theta_*}^{\theta_H} \left\{ x(\theta)\pi_M(\theta) + (1 - x(\theta))\left[\pi_1(\theta) + \pi_2(\theta)\right] - I - \int_{\theta_*}^{\theta} x(t)\left[\dot{\pi}_M(t) - \dot{\pi}_1(t)\right] dt \right\} dF(\theta) \geq 0,$$

and after integration by parts,

$$\int_{\theta_*}^{\theta_H} \left\{ x(\theta)\left[\pi_M(\theta) - \frac{1 - F(\theta)}{f(\theta)}\left[\dot{\pi}_M(\theta) - \dot{\pi}_1(\theta)\right]\right] + (1 - x(\theta))\left[\pi_1(\theta) + \pi_2(\theta)\right] - I \right\} dF(\theta)$$

$$\equiv \int_{\theta_*}^{\theta_H} \left\{x(\theta)B_M(\theta) + (1 - x(\theta))B_D(\theta)\right\} dF(\theta) \geq 0. \quad (16)$$

The program that determines the optimal policy is then:

$$\max_{x(.)} \int_{\theta_*}^{\theta_H} \left\{x(\theta)W_M(\theta) + (1 - x(\theta))W_D(\theta)\right\} dF(\theta)$$

s.t.(16) and $x(.)$ non-decreasing.

Letting μ denote the Lagrange multiplier associated to the constraint, the Lagrangian can be written as:

$$\int_{\theta_*}^{\theta_H} \{x(\theta)\,[W_M(\theta) + \mu B_M(\theta)] + (1 - x(\theta)\,[W_D(\theta) + \mu B_D(\theta)]\}\, dF(\theta).$$

The unconstrained optimization of this Lagrangian would deliver a bang-bang solution in terms of x, that is $x(\theta)$ would be either equal to 0 or 1. Optimizing with the monotonicity constraint, however, imposes $\dot{x} \geq 0$. If the solution is such that on an open set $\dot{x} > 0$, then this solution should locally be the same as the solution to the unconstrained program, which can never be strictly upward sloping.

Consequently, we obtain the following claim:

Claim 11 *: The optimal stochastic policy $x(.)$ is a step (increasing) function, with bunching on subintervals of $[\theta_*, \theta_H]$.*

That is, eliciting information from the incumbent is of little help in designing the optimal policy. The complete derivation of the solution follows standard optimal control methods (see e.g. Guesnerie-Laffont [1984]).

Let us finally turn to the specific program in the text with a constant x. The limit values for x determine the value of the corresponding θ_*, namely: $x = 0$ implies $\theta_* = \theta_D$ while $x = 1$ implies $\theta_* = \theta_M$. The constraint shows that x and θ_* are linked so that:

$$\frac{d\theta_*}{dx} \mid_{x=0} = -\frac{\pi_M(\theta_D) - I}{\dot{a}_D(\theta_D)} < 0,$$
$$\frac{d\theta_*}{dx} \mid_{x=1} = -\frac{I - a_D(\theta_M)}{\dot{\pi}_M(\theta_M)} < 0.$$

The total derivative of social objectives w.r.t. x taking the constraint into account, is given at $x = 0$ by:

$$-\int_{\theta_D}^{\theta_H} [W_D(\theta) - W_M(\theta)]\, dF(\theta) + W_D(\theta_D) f(\theta_D) \frac{\pi_M(\theta_D) - I}{\dot{a}_D(\theta_D)}.$$

The first term is negative; the second term is positive and proportional to $f(\theta_D)$. If we let the density $f(.)$ increase on a left neighborhood of θ_D, the positive effect of expanding the range of implemented projects dominates the cost of granting a monopoly franchise with some probability on all inframarginal projects. So, $x > 0$ at the optimum.

Similarly, the total derivative of social objectives w.r.t. x at $x = 1$ is given by:

$$-\int_{\theta_M}^{\theta_H} [W_D(\theta) - W_M(\theta)]\, dF(\theta) + W_M(\theta_M) f(\theta_M) \frac{I - a_D(\theta_M)}{\dot{\pi}_M(\theta_M)},$$

which can be made negative by letting the density $f(.)$ be small enough in a right neighborhood of θ_M. So, $x < 1$ at the optimum.

Item ii) of the proposition follows.

B Appendix 2: Proof of Proposition 7

Letting λ denote the shadow price of the network-wide budget constraint, the derivatives of the Lagrangian L with respect to $\hat{\theta}_M$ and $\hat{\theta}_D$ are:

$$\frac{\partial L}{\partial \hat{\theta}_M} = m_M f(\hat{\theta}_M) \left[-W_M(\hat{\theta}_M) + \lambda \left[\frac{1 - F(\hat{\theta}_M)}{f(\hat{\theta}_M)} \right] \dot{\pi}_M(\hat{\theta}_M) + \lambda \left[I - \pi_M(\hat{\theta}_M) \right] \right]$$

and

$$\frac{\partial L}{\partial \hat{\theta}_D} = m_D f(\hat{\theta}_D) \left[-W_D(\hat{\theta}_D) + \lambda \left[\frac{1 - F(\hat{\theta}_D)}{f(\hat{\theta}_D)} \right] \dot{a}_D(\hat{\theta}_D) + \lambda \left[I - a_D(\hat{\theta}_D) \right] \right].$$

Note that the sign of these derivatives does not depend directly on the proportion of segments that are assigned to either regime.

We now show that, under assumption A.6, open access segments are subsidized by the monopoly segments. Suppose to the contrary that

$$a_D(\hat{\theta}_D) - I \geq 0 \geq \pi_M(\hat{\theta}_M) - I$$

(from (11), these net contributions to the network's budget must have opposite signs). Then from A.5, $\hat{\theta}_D > \hat{\theta}_M$. The monotone hazard rate assumption implies that

$$\frac{1-F(\hat{\theta}_D)}{f(\hat{\theta}_D)} < \frac{1-F(\hat{\theta}_M)}{f(\hat{\theta}_M)}.$$

Assumption A.6, together with $\hat{\theta}_D > \hat{\theta}_M$, then implies that $\partial L/\partial\hat{\theta}_M = 0$ and $\partial L/\partial\hat{\theta}_D = 0$ are inconsistent.

Last, let us show that it may be optimal to treat otherwise identical segments differently. To this purpose, consider first the special case in which open access and monopoly franchise are equivalent in the single-segment context (that is, (6) is satisfied with equality). Then social welfare is insensitive to the number of segments operated under open access as long as $\hat{\theta}_M = \theta_M$ and $\hat{\theta}_D = \theta_D$ (recall that $a_D(\theta_D) - I = 0 = \pi_M(\theta_M) - I$). From the previous reasoning, welfare can be strictly improved by lowering $\hat{\theta}_D$ below θ_D and increasing $\hat{\theta}_M$ above θ_M. By continuity, the same holds when (6) is close to be satisfied with equality.

Charging for the Use of Railway Capacity

Stephen Gibson
Regulatory Economics Manager, Railtrack

Background

One element periodic review of Railtrack's access charges for the period 2001-2006 is to review and improve the structure of access charges. The aim is to construct an efficient structure of charges that recovers the (short run) marginal costs of track access through variable charges (asset usage, electric traction charges and congestion charges) plus a volume incentive to incentivise growth in the use of the network. There is also a fixed charge to recover Railtrack's residual revenue requirement over the 5 year period. There is also a set of performance regimes to incentivise reliability (delays and cancellations) on the network.

Performance Regimes

Railtrack has a performance regime with each train operator. Under these regimes, Railtrack is responsible for all delay/cancellation not directly caused by the operator including reactionary delay (when operator A's trains delay Operator B).

Railtrack compensates operators for delay/cancellations above historic benchmarks, while operators reward Railtrack if performance is better than the benchmarks. The payment rates are based on the value/cost of the delay to the operator and vary between different operators and for an operator by service group[1].

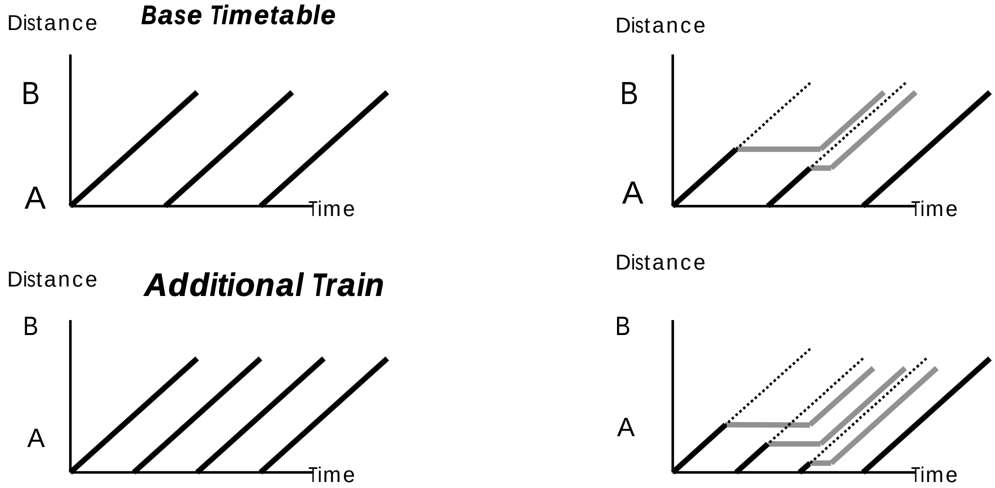
Congestion Costs

Congestion costs arise as a result of an additional train increasing the reactionary delay on the network (even if that train operates on time) and therefore imposing a cost on Railtrack through the performance regimes. The diagram below demonstrates this increase in reactionary

[1] Operators divide their services into 1 to 6 service groups

delay through comparing the effect of an exogenous delay on a timetable containing 3 trains per hour with the effect on a timetable containing 4 trains per hour. It is clear that the more congested timetable has a higher level of reactionary delay, and since Railtrack compensates operators for this delay through the performance regimes, Railtrack will face a marginal congestion cost associated with timetabling an additional train.

Effect of Exogenous Delay on Reactionary Delay



Current recovery of Congestion Costs

The existing access charges recover the **average** congestion costs for the **initial** set of rights from the fixed charge but include no variable charge to reflect changes in the use of those initial rights which change the congestion costs they give rise to.

The expected congestion costs for **additional** rights are calculated and negotiated on a case-by-case basis as the requirement for those rights arises.

Capacity Charges

Railtrack proposes the introduction of a tariff (published in advance) for initial and additional rights on the existing network. It would be paid when those paths are bid into the timetable (when the capacity is used) and would be known as the 'Capacity Charge'.

The capacity charge could have up to 4 dimensions, reflecting the underlying cost drivers: location; time of day/day of week; relative speed and flex in pathing the access right.

Benefits of a Capacity Charge

A capacity charge would incentivise the best use of scare capacity and reduce overall delay. It would signal (but not fund) where investment in additional capacity is likely to offer value for money. It would also make charges transparent and facilitate forward planning, and would avoid the transactions costs of negotiation.

Defining Capacity Utilisation

To measure and therefore charge for congestion, a consistent measure of network capacity utilisation was defined. This takes the timetabled trains currently running in a set time period (normally an hour) and ‘squeezes’ them together (keeping the order constant) until the trains are at the minimum headway apart. The time it is possible to run these ‘squeezed’ trains in, as a proportion of the actual timetabled time used, is defined as the capacity utilisation index.

Capacity Utilisation and Delay

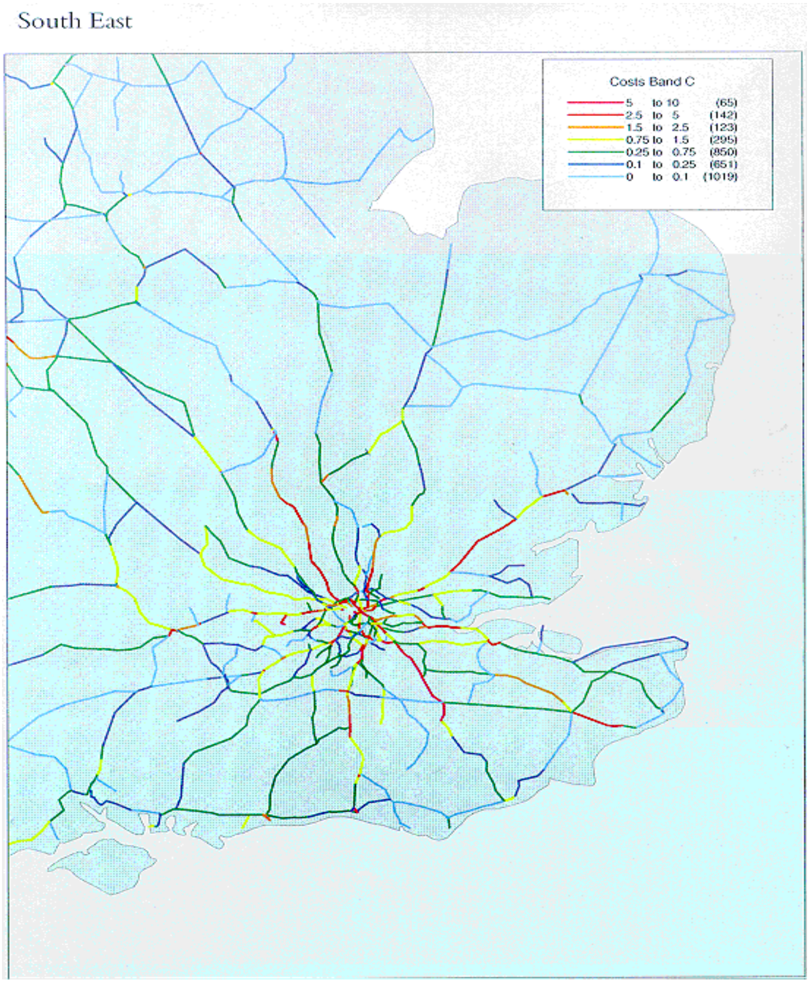
Railtrack used multiple regressions of data on delay and capacity utilisation across the network to establish an exponential relationship of the form.

$D = A \exp(ß C)$

Where D = delay per train
A = section constant
C = capacity utilisation index of the route
ß = route constant

Distribution of Congestion Costs

Congestion costs vary significantly by time and location across the network as shown in the map below:



Constructing a Capacity Charge Tariff

A compromise had to be reached between complexity and accuracy to ensure that the capacity charge tariff was sufficiently granular to pick up significant variations in costs across the network, while sufficiently simple to be practical to implement. The decision reached was to introduce a tariff with 1700 geographic cells (1350 route sections in each of 2 directions) and 13 timebands (7 for weekdays, 4 for Saturdays and 2 for Sundays). The complexity of the charge was reduced significantly by banding the costs into 10p/train mile charging bands and introducing a *de minimis* level of 10p/train mile (which covers about 33% of the charging cells) below which charges are set to zero.

Capacity Charge

An example of the congestion costs for the Midland Mainline network is show below:

MIDLAND MAINLINE: Sheffield - London St. Pancras

Congestion costs (pence per route section)				
		Weekday - Timeband		
From	To	0630-0930	0930-1630	1630-1930
Sheffield	Dore Station Jn	92	82	64
Dore Station Jn	Chesterfield	53	44	30
Chesterfield	Clay Cross South Jn	29	26	20
Clay Cross South Jn	Ambergate Jn	54	66	37
Ambergate Jn	Derby	74	54	53
Derby	Long Eaton	82	29	35
Long Eaton	Syston South Jn	49	38	39
Syston South Jn	Wigston North Jn	78	49	57
Wigston North Jn	Bedford	25	16	23
Bedford	Luton	462	19	24
Luton	St Albans	461	34	33
St Albans	Kentish Town	486	126	64
Kentish Town	London St Pancras	18	10	22

These costs would then be banded into 10p/ per train mile charging bands and the de minimis threshold applied in order to construct the capacity charge tariff.

Conclusion

Railtrack have developed a tariff to reflect the marginal costs of using capacity on the UK rail network. It will be introduced on the UK network from April 2002. It will be interesting to see how timetabling behaviour is influenced by a charging mechanism that explicitly recognises the congestion costs of operating trains on the network.

TWO-PART TRACK ACCESS CHARGES IN GERMANY: THE CONFLICT BETWEEN EFFICIENCY AND COMPETITIONAL DISCRIMINATION

Paper prepared for the European Seminar on Infrastructure Charging

Helsinki, 31 July - 1 August

By

Heike Link[1]

German Institute for Economic Research, Germany

[1] Dr. Heike Link, German Institute for Economic Research (DIW), Department of Energy, Transport and Environment, Königin-Luise-Str. 5, 14 195 Berlin, Germany

phone: ++ 49 - 30 - 897 89 312, fax: ++ 49 - 30 - 897 89 103, e-mail: hlink@diw.de

1 Introduction

In 1994 the rail network of Deutsche Bahn (DB) was opened up for third parties against payment of access fees. Although meanwhile a number of non-DB companies operate services at DB-tracks the share of train-km they run there is still low. The obvious reason for this situation is the institutional framework which is not appropriate for the German rail market: It does not imply any regulation on the incumbent company DB which provides the tracks both for own services and for the services of the competitors.

This broad and obvious diagnosis of the problems in the German rail market is not sufficient to derive more detailed policy recommendations. However, a more sound and deep analysis is faced with the fact that almost no information is available on the number of non-DB users of the network, the exact number of passenger- and freight-km they run and on the problems they face in the context of network access and access charges. This was the reason to carry out an empirical study aimed at gaining deeper knowledge into the competitional situation of the German rail market. This paper reports the results of this empirical research, a company survey with DB-competitors.

The paper is organised as follows: Chapter 2 summarises the philosophy and the main features of the German railway reform of 1994. Theoretical foundations for infrastructure charging are summarised in chapter 3. The practice of network access and access charging in Germany are detailed in chapter 4. Chapter 5 describes the findings of the company survey and the survey results. Chapter 5 concludes.

2 Main features of the German Railway reform from 1994

In 1994 a fundamental reform of the German national rail companies Deutsche Bundesbahn (DB) and Deutsche Reichsbahn (DR) was introduced which was designed as a 10 years process with the following measures:

- The foundation of Deutsche Bahn AG (DB AG) as a private sector company in 1994,
- The institutional separation between infrastructure and transport,
- Opening up of the rail network for third parties against the payment of track charges,
- Federal responsibility for rail infrastructure investments,
- Financial refloating measures on the part of the state,
- Regionalisation of suburban passenger transport from 1996 onwards.

Most important for the topic of this paper are two main features of the general restructuring philosophy, namely (i) to separate the network from the operational side institutionally, and (ii) to open up the network for third parties in order to attract more rail traffic. However, despite of remarkable steps towards this philosophy, the ongoing process is to some extent inconsistent.

To start with, the institutional separation between infrastructure and transport services was in a first stage (from 1994 up to 1998) realised by splitting up DB AG into four subdivisions (tracks network, long-distance passenger transport, regional passenger transport, freight transport). In a second phase (starting with 1999) these subdivisions became public limited companies in their own right which, however, operate under the roof of Deutsche Bahn Holding. A third phase is only optional and would - if realised - imply to dissolve the Holding. This third phase has not an obligatory character: As a consequence, the DB group with four companies including the track provider under one roof has still all features of a vertically integrated company. If the third phase will not be introduced which seems to be the current preference of the German transport policy, these features will remain.

Further problems, closely connected with the existence of a vertically integrated incumbent, arise with network access and access charging rules. Although the introduction of competition was not an explicit aim of the reform, DB opened up the rail network for third parties against the payment of access charges for the use of infrastructure already in the middle of 1994. Germany belongs together with the U.K. to the most advanced countries in implementing the EC-directive 91/440 EWG. However, apart from some general rules for network access and access charges, there does not exist any regulation of the incumbent, which would be the necessary precondition for new entrants to compete with the incumbent, as more as the incumbent is a vertically integrated monopolist.

3 Theoretical remarks on Rail track access charging

The problem

The problem of track access charging by a vertically integrated incumbent to its competitors is characterised by the following issues:

- The track network is a natural monopoly.
- The competitors (entrants) and the incumbent produce the same final product and need network access for the production process.
- In rail transport – in contrast to other utilities such as telecommunication, - no bypass is possible. However, final consumers have an inter-modal choice[2].
- The track network is characterised by high fixed costs which imply decreasing average costs. Marginal costs are below average costs which means that the 1st best pricing rule of charging marginal costs leads to a financial deficit of the track provider.

From these characteristics several competitional problems arise: On the one hand, the charging policy of the incumbent can prevent market entry by setting too high prices. On the other hand, too low access charges might lead to inefficient market entry and cream-

[2] This argument is often used to state that rail transport is actually a market with competition from other transport modes and, consequently, competition in rail transport itself is not necessary.

skimming. Charging of access to input factors which are provided by an incumbent company to its competitors and which are at the same time needed by this incumbent company itself for producing its services is therefore one of the most complicated economic problems.

Theoretical pricing principles

Depending on the constraints to be considered economic theory has provided a range of possible pricing rules.

The classical 1^{st} best pricing rule of charging marginal costs was chosen by some European countries (for example Sweden) as a basis for network access charges although on has to bear in mind that due to lacking knowledge on cost functions the practically applied access charges in these countries provide rather approaches to marginal costs than genuine marginal cost pricing. If budget constraints are binding, e.g. if the infrastructure provider is expected to cover the fixed costs of the network (or at least a portion of fixed costs) the price = marginal cost rule does not work. For these cases economic theory has provided several approaches deviating from marginal cost pricing: One of the historically oldest was developed by Ramsey 1927 and Boiteux 1971 who derived pricing rules containing a mark-up over marginal costs by the inverse price elasticities of demand. The problem of these approaches consists in the lack of data and knowledge on price elasticities.

The regulation practice in the US has developed several pricing rules which can be summarised under the term “fully distributed costs”. These approaches are average cost principles and use a variety of methods to allocate fixed, common and joint costs to users. The problem of these approaches is that this allocation tends to be more or less arbitrary since common and joint costs are per definiton not allocatable to groups of users (for a critical discussion see Braeutigam 1980.The critics to these pricing rules led to the development of the concept of stand-alone costs according to which only the costs which a service causes alone are allocated (as it were isolated from other services together with which joint and common costs are shared).

Another stream of theory has started with Willig 1979 and Baumol 1983 who developed the so-called efficient component pricing rule. This rule says that the access charge should contain the direct costs of providing network access to users and the opportunity costs of this provision occurring for the incumbent in form of lost profits by not using the network himself. More recent work by Laffont and Tirole 1994 and Armstrong et al. 1996 has started with this pricing rule and has developed this rule further by stepwise relaxing of the underlying assumptions. They yield more complicated, although linear pricing rules, consisting of a number of additive price elements such as marginal costs, Ramsey terms, displacement ratios etc.

All the pricing rules mentioned so far have in common that they are – more or less complicated – linear pricing rules[3]. Already in the late 70es another stream of theoretical

---

[3] This linearity, however, should not be confused with simplicity. In fact many of these pricing schemes are derived from sophisticated micro-economic theory and contain terms which are difficult to estimate both with respect to data and eeconometric estimation procedures.

literature has proposed non-linear tariffs for charging utilities, with self-selection of the linear or non-linear tariff by the final consumer.-. This literature deals, however, with final products. The theoretical foundations for non-linear tariffs with self-selection were elaborated by Feldstein 1972, Faulhaber and Panzar 1976, Spence 1977 and Willig 1978. Self-selection means that purchasers of the commodity in question can select between a linear (one-part) tariff and a non-linear two-part tariff consisting of a fixed entrance fee and a quantity-dependent variable charge. Willig 1978 has shown that non-linear outlay schedules with self-selection are under certain conditions pareto-superior compared with linear tariffs. The conditions to be met for this optimality feature of non-lineer tariffs with self-selection are:

i) The total quantity purchased by each consumer can be monitored by the vendor.
ii) The commodity cannot be traded among the purchaser.
iii) The marginal price for the largest purchaser must equal marginal costs, otherwise the nonliner outlay schedule can be dominated by another non-linear outlay schedule.

It is important to mention that the theoretical literature on non-linear outlay schedules has dealt with utilities, but only with final products. In contrast to that, network access is an input needed both by the incumbent and by the competitors for producing transport services (e.g. the final products).

Due to the pareto-dominance of non-linear outlay schedules the introduction of this tariff type for network access charging of DB was welcomed by national transport economists[4]. Also the Antitrust-commission did not see any serious problems in the beginning the introduction. However, the pre-conditions mentioned above for the optimality of this tariff-form are violated. This concerns first of all the fact that the pareto-dominance has only been proven for final products, not for the access charging problem of a vertically integrated monopolist. Furthermore, the condition that the marginal price for the largest consumer has to equal marginal track costs is not met. Therefore, we have here the situation that theory was not properly applied to practice. Apart from this, the practical competitional problems have meanwhile convinced the Antitrust-Commission to become active. The Commission has obliged DB Netz to elaborate a new pricing scheme by next year. As we will see in this paper also the empirical results on the market situation support this decision.

4 The Current situation of Network access and access charges

4.1 Rules for Network Access

Germany has realised the most comprehensive opening up of a rail network in Europe. Since 1994, DB AG and all other rail companies offering public rail transport services

[4] Cf. Aberle 1998 and Knieps 1998.

have opened up their routes against the payment of usage charges to the following groups of users which exceed those defined in the EC-regulation 91/440[5]:

- public-transport railway companies which possess an own rail network,
- non-public-transport railway companies who likewise grant - under similiar terms - other public railway companies access to their infrastructure,
- railway companies from EU countries for cross-border intermodal traffic,
- foreign railway companies, if mutual access to the rail network is guaranteed, otherwise on the basis of international agreements.

DB AG even grants - in addition to the mentioned groups - other companies such as haulage contractors, travel companies and government bodies access to its routes. Responsible for the operation and management of the tracks and the negotiations with companies applying for network access is DB AG's track company.

After a long period where any regulatory framework did not exist[6], a regulation on the use of rail infrastructure was passed. This regulation contains the following rules:

- All companies providing rail infrastructure for third parties (this concerns apart from DB AG all transport companies offering public rail services) are allowed to define the level and structure of track charges freely.
- Rail infrastructure managers must not apply for approval of track charges. There does not exist any regulation of prices by an independent authority.
- It is admissable to charge average prices for the network as a whole, or to calculate charges for parts of the networks or certain routes.
- Factors to be considered in the definition of track charges can be: Types of routes, operating days and times for the services, operated vehicle types, wear-and-tear of the tracks, utilisation of tracks. Track access charges may also depend on air and noise pollutions of trains.
- The track provider is not obliged to publish the charge system but has to provide insight into the system if customers request this insight.
- The track companies can provide quantity discounts for ordering a certain volume of train-kilometres. However, they have to prove (by an official certificate of an auditor) that this is justified by respective cost reductions.[7]
- In case of competing demand for tracks the track companies are allowed (but not obliged) to grant track access to the most-bidding customer.

As stated in these rules, there is no price regulation for the access charges. If competitors of DB feel discriminated they have the chance to claim

---

[5] The regulation claims only access for rail-companies from EU countries in cross-border intermodal traffic.

[6] DB AG's network company was completely free to design /redesign the structure of track charges and to define their level. The mechanisms to control the non-discriminatory network access did either not exist or were insufficient. Also the problem of granting linear increasing discounts for ordering certain amounts of tracks which clearly favoured the transport companies of DB AG as the most important customers of track company was not solved.

[7] So far there is no case known that this proof was requested and provided.

i) against the antitrust commission,
ii) the federal railway office (Eisenbahnbundesamt - EBA).

This means that there are two institutions responsible for enabling a fair level playing field for the incumbent and the competitors. Both institutions have only passive regulatory power, e.g. they can only react on the basis of received claims but are not entitled to active market regulation. While the antitrust commission is responsible for all claims concerning competitional questions on the basis of the German Act of restraints against competition[8], EBA deals mainly with issues of technical system's compatability, safety, qualification requirements for drivers etc. Interesting in regard to track access charges, however, is the fact that competitors can claim against both institutions. The antitrust commission has the possibility to introduce a formal procedure to prohibit the price system, while EBA can according to § 14 (5) AEG[9] even decide on the level of prices in the claimed case.

4.2 Access charges

4.2.1 Description of the tariff system

Charging for the use of rail tracks and station has meanwhile a 6-years history in Germany. The first price system for track use was introduced in 1994, followed by a price list for the use of stations in 1995. In the mean-time there were already two revisions of the track charge system, the first one in 1995 and a second, more fundamental one in 1998. All track charge systems have so far the average cost principle applied. The first two charging regimes were based on a linear tariff with quantity discounts. The system valid since 1998 is a non-linear tariff with self-selection. Due to the anti-competitive features of this current version of access fees the Anti-Trust Commission has obliged DB to a further revision of the system which is expected by 2001. The following descriptions and the survey results refer, if no other explanations are made, to the charging regime valid from 1998 to 2000.

The track fees are charged for providing the tracks (including passing- and crossing tracks and the tracks within stations) as well as for operating the network and compiling the time table. The use of stations, marshalling yards etc. has to be paid by special access charges defined in an additional price framework mentioned above. Furthermore, the track charges do not include the consumed electric power for the electrified sections, shunting services and VAT.

While the track access charges during the period from 1994 up to 1998 were based on a one-part tariff[10], the current charging system follows a non-linear outlay schedule with

---

[8] Cf. Gesetz gegen Wettbewerbsbeschränkungen. Bundesgesetzblatt 1998, Teil I, Nr. 59, S. 2546.
[9] AEG: Allgemeines Eisenbahngesetz (Railway Act). Bundesgesetzblatt 1993, Teil I, S. 2396.
[10] The one-part tariff, introduced in 1994, consisted (i) of basic prices, (ii) of modification factors according to user's requirements (reliability, special types of operated trains), and (iii) of quantity and contract duration. Due to heavy critics of these prices regarding i) the high level, in particular for regional passenger trains, ii) the discriminatory discounts favouring the

self-selection. Customers of the track company can choose between a two-part (non-linear) tariff and a one-part (linear) tariff which is, however, compared with the variable part of the two-part tariff, considerably higher. For both types of tariffs the price differentiation according to train types (which was an element of the former tariff systems) was abolished. In the current access charging regime the price depends only on the route type (differentiated into 6 route classes, compared with 10 classes in the former systems) and the level of utilisation on these routes (with three route classes regarding the utilisation). The fixed entrance fee depends furthermore on the type of traffic (regional passenger transport, long-distance passenger transport, freight transport). Discounts are only granted for long-term commitments of using tracks. Modifications of the prices are made for special train dimensions and in particular for the degree of time-table flexibility. For the latter, a maximum discount of 50% can be granted in case of complete time-table flexibility. Possible options for future considerations are reduced prices in order to support innovative train systems and for low-noise trains.

The non-linear two-part tariff consists of a fixed charge (so-called InfraCard) and a usage related part depending on train-km. The fixed part is charged for certain routes which the user can select. Precondition is, however, that these routes form a network, e.g. the two-part tariff cannot be applied for a set of single lines. The possibility to select a network was introduced in order to guarantee more fairness in the contribution to regional differences in network costs. A further precondition for applying the two-part tariff is to serve a minimum network length which was originally defined to be 100 km in regional rail passenger transport, 1000 km in long-distance passenger transport and 500 km in freight. For those users which (i) do not meet this minimum requirement, or which (ii) have a too low train frequency, the one-part tariff is applied which consists only of a usage related component (so-called VarioPrice).

There were critics from new entrants that the minimum length required to be served would built up entry barriers for new entrants, since in such cases the one-part tariff, only depending on train-km, with much higher price levels, would be applied. Resulting from this debate, the minimum network length for the two-part tariff was reduced to 25 km in regional passenger services, 800 km for long-distance passenger services and to 240 km for freight services.

4.2.2 Comparison of one-part and two-part tariff access charges

The main difference between the linear one-part tariff (VarioPrice) and the non-linear two-part tariff consists in the shape of the price curves per production unit, e.g. per train-km. As figure 1 shows, the two-part tariff has a degressive effect with increasing train-km while the one-part tariff gives a constant amount to be paid for each train-km.

---

incumbent and iii) the assumed cross-subsidisation of DB's freight business by track revenues from regional passenger transport10, the price system was already revised by January 1995 (amongst others: limitation of possible discounts, reduction of the price level for regional rail passenger transport).

Figure 1
The shape of track access charges per train-km for one-part and two-part tariff

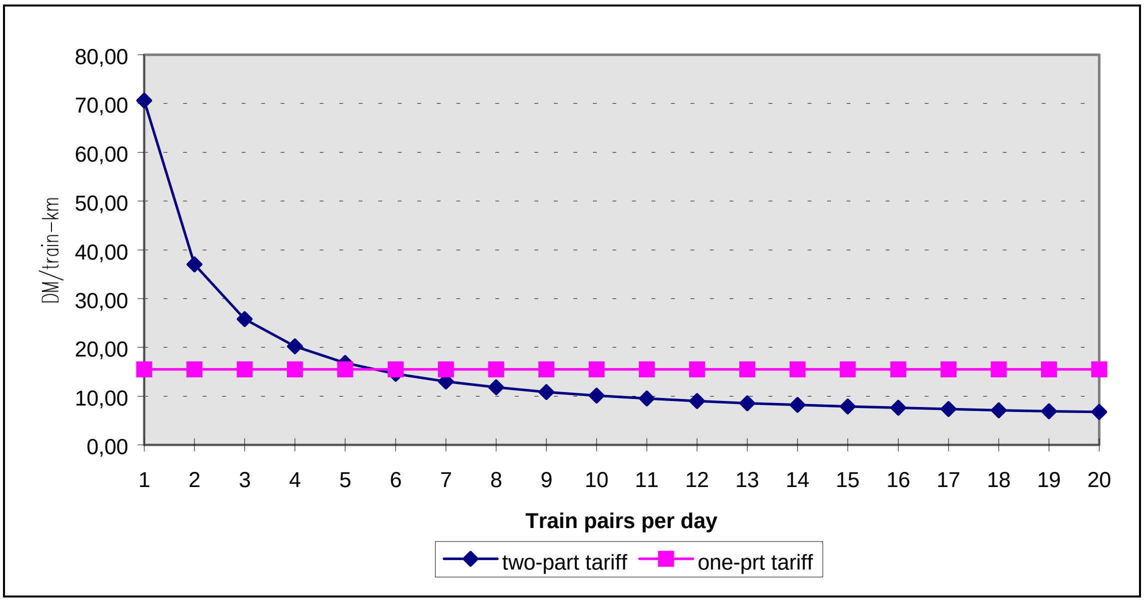


Figure 2
Access charges in regional passenger transport for a 25 km line (minimum length for two-part tariff) at route type K1- in DM mill.

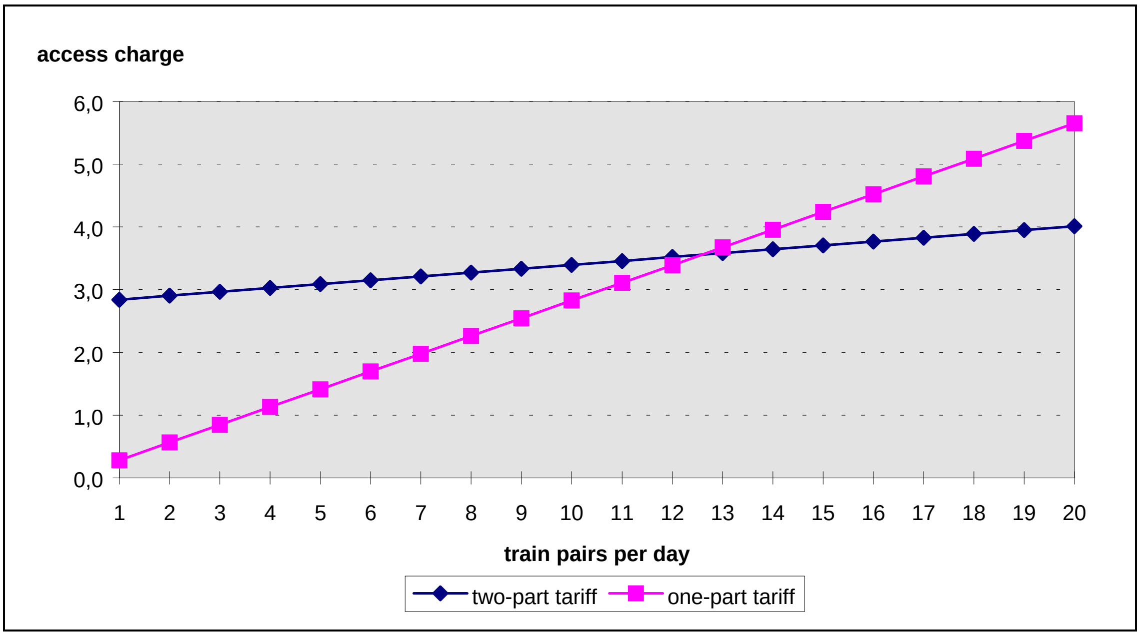


Figure 2 compares the prices resulting from the one-part and the two-part tariff for regional passenger transport. The curves refer to the minimum network length to be served for the two-part tariff (25 km) transferred into train-pairs per day. Figure 2 shows the general problem which new entrants are faced with: In a situation of 20 regional passenger train pairs per day the prices resulting from the one-part tariff exceed those from the two-part tariff by 41%. Only very few competitors are probably able to reach such traffic levels to break even between one-part tariff and the two-part tariff. Of course, also competitors of DB are able to profit from these price differences between one-part and two-part tariff, but due to the fact that the DB companies use more tracks than the competitors, DB can profit even in case of low frequencies on single lines from the InfraCard paid for a larger network.

5 The competitional situation at the German rail market - results from a company survey

The information policy of DB with respect to the degree of competition on DB's tracks was so far rather scanty or simply non-existent. DB only released information on the number of contracts agreed with third parties on network access. According to this information released, about 150 contracts were signed. However, it is not clear to which period this, whether the same company using tracks in passenger and freight transport is counted once or twice etc. For the years 1994 - 1998 there do not exist any figures on the train-km run by third parties on DB tracks and the revenues collected through access fees from them. The competitional which non-DB companies face when using DB' tracks are reported in the daily press and it is known that 3 competitors have claimed against the antitrust Commission. This brief characterisation shows that there was a clear need for gaining more insight into the situation. DIW has therefore carried a company survey which addressed the competitional problems of DB competitors. The survey design and the survey results are described in the following sections.

5.1 Survey design

The survey design had first of all to consider that the number of potential users of DB-tracks is considerable and in fact unknown due to the comprehensive opening up of the network. Therefore, the sample was designed to comprise both rail companies and rail-bound public transport operators and it was decided to restrict the survey to domestic companies. Companies which run both passenger and freight transport business were treated as two separate users for the following reasons:

i) They have to buy more than one InfraCard.
ii) They have to sign separate contracts with DB.
iii) The survey was aimed at treating the problems of freight and for passenger transport separately.

This means that the discussion of results refers not to the number of companies but to the number of track users.

The survey was designed as a mail survey with a comprehensive questionnaire containing four parts:

– Part A: Situation of competitors (use of DB-network, access problems, access charging)
– Part B: Evaluation of the competitional situation (in particular with respect to access rules, access charging, information on access rules, price-relevant factors, charges etc., institutional framework, problems with vehicle purchase/rental and staff)
– Part C: Suggestions and wishes of competitors
– Part D: Company information

Table 1: Sample description

	Total	Passenger transport	Freight transport
1. Questionnaires	331	85	246
2. Responses	131 (39%)	47 (53%)	84 (34%)
3. Valid answers for analysis	55	27	28
out of these: users of DB-network	47	25	22
4. Companies with own network	35	14	21
size of own network:			
less than 20 km	17	7	7
21 - 50 km	5	1	4
51 - 100 km	8	3	5
101 - 200 km	6	2	4
more than 200 km	2	1	1
5. Number of employees			
less than 25	22	9	13
26 - 50	10	5	5
51 - 100	12	8	4
more than 100	11	5	6
6. Companies with own vehicles			
Locomotives	46	20	26
Waggons	15	11	2
Source: DIW.			

As table 1 shows the survey achieved with 40% a good response rate. However, more than half of the responses received stem from companies which do not use DB's network and which therefore did not fill in the questionnaire. After correction by unplausible answers 55 answers were used for the analysis. These 55 answers contain users of DB-tracks and potential users which explored the opportunities to use DB-tracks, which eventually. already faced discrimination etc.

The survey yielded a number of 43 non-DB users of DB’s track. Out of these, 21 companies operate passenger services (mainly regional services), 18 companies run freight services and 4 companies operate both types of services. A separate count of companies being involved both in passenger and freight transport, as mentioned above, yields therefore 47 cases.

Most of the non-DB companies using DB’ tracks possess also an own rail network (two thirds of the users in passenger transport and more than three quarter of those in freight transport). With respect to the company size we can state that in case of companies using DB-tracks for passenger transport the distribution regarding the number of employees is rather equal: companies with employees less than 25 and between 26 and 100 have both a share of 38%, one quarter employs more than 100. The corresponding figures for those companies using DB tracks for freight transport (48%, 22% and 30%) show a larger share of small companies. Most companies possess own locomotives (46 cases out of 55), however, only one quarter of DB-competitors have waggons in their property. As the survey results show (section 4.4.5) this causes considerable problems for them.

5.2 The track access fees charged to DB-competitors

DB-competitors use only in 7 cases the two-part tariff (see figure 3), exclusively in regional passenger transport. The main reasons for this low number are that mostly the competitors are not able to reach the break-even point between one-part and two-part tariff and that they often do not meet the requirement of serving a network of a minimum length. This shows clearly that the track access charging regime in Germany is discriminatory against DB-competitors. In addition, only one quarter of the interviewees answered that the two-part tariff gives them incentives to increase traffic volume on the network. This shows that the argument of the supporters of this tariff system, namely the ability to attract more traffic, does not proof to be a real one.

It is not surprising that the competitors evaluated the current track access charging system as poor (see figure 4). Consequently, only in two cases the DB-competitors favoured to keep the current system. However, obviously the competitors also disfavoured a return to the former one-part tariff even if it would include – in contrast to the former system – only restricted discounts. This might be explained by the bad experience they have made with this former tariff system. However, it shows also that to suggest a new charging system is a complicated and complex issue. More than 85% of the competitors have expressed their wish of a general reform of the tariff system or the introduction of the so-called virtual InfraCard.

Figure 3
Reasons mentioned by DB-competitors for using the one-part tariffs (multiple answers)

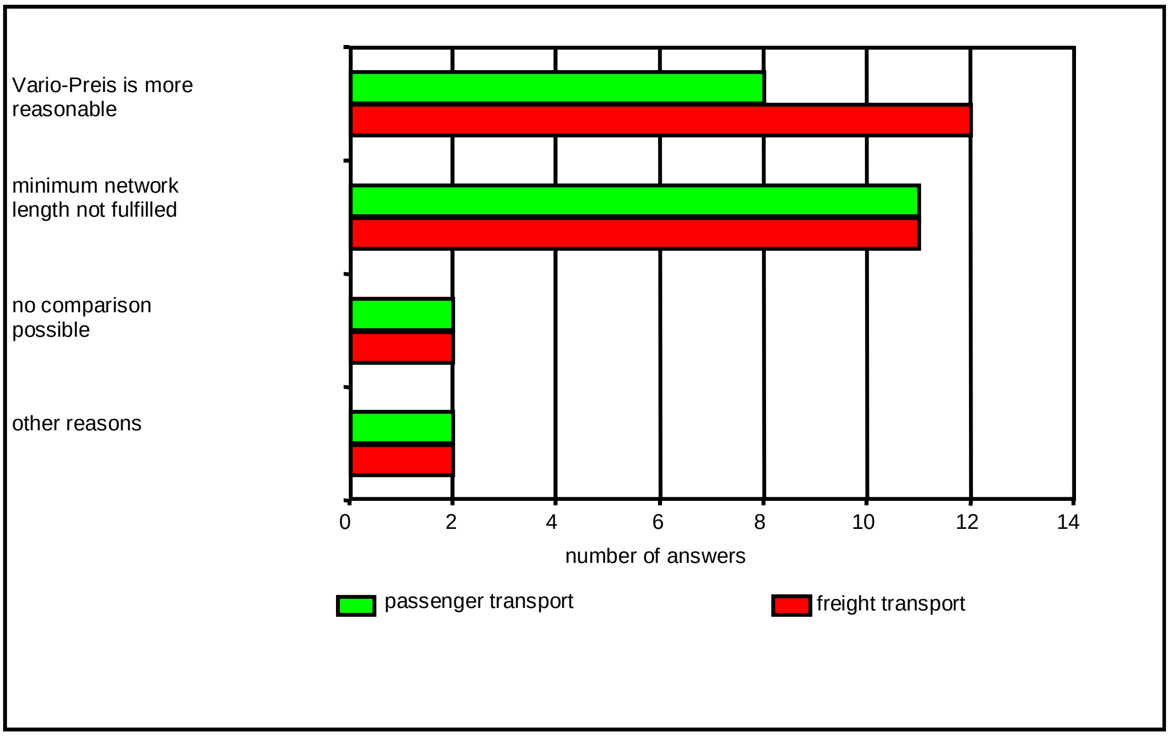
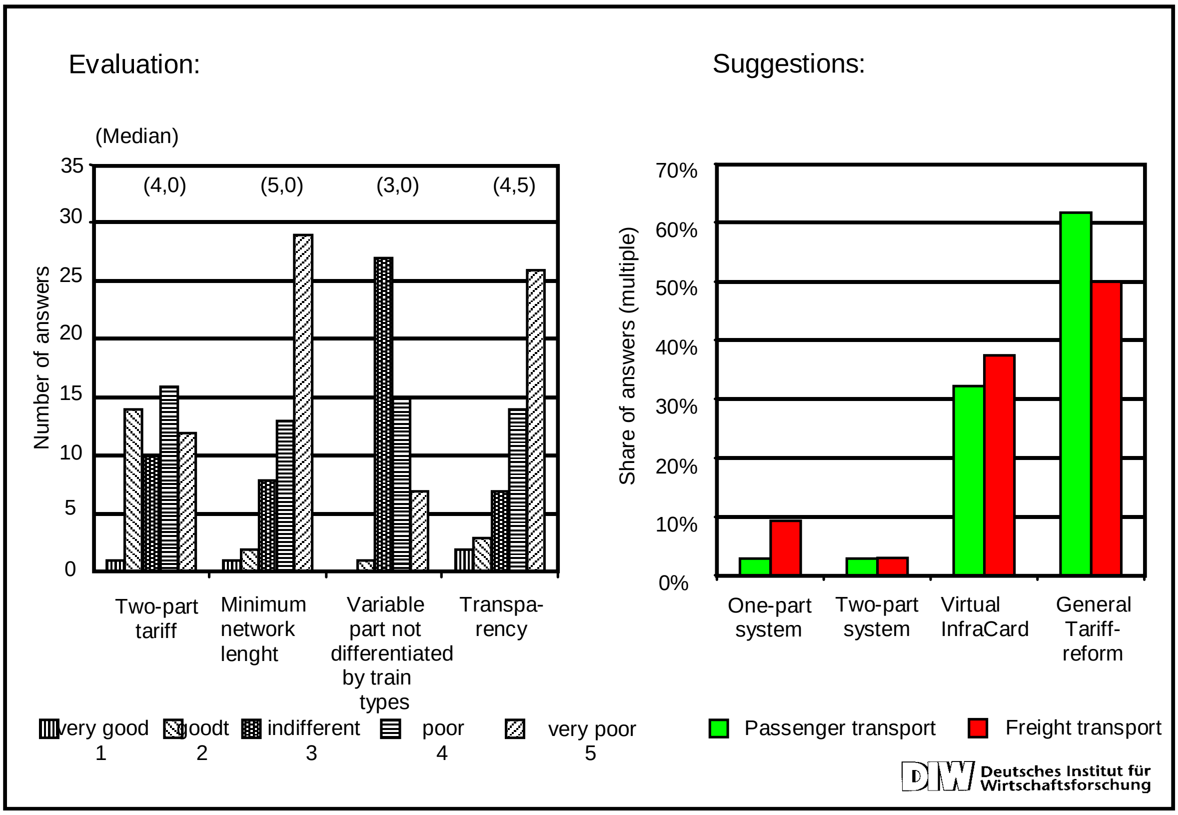


Figure 4
Evaluation of DB’s track access charging regime and suggestions for reforms

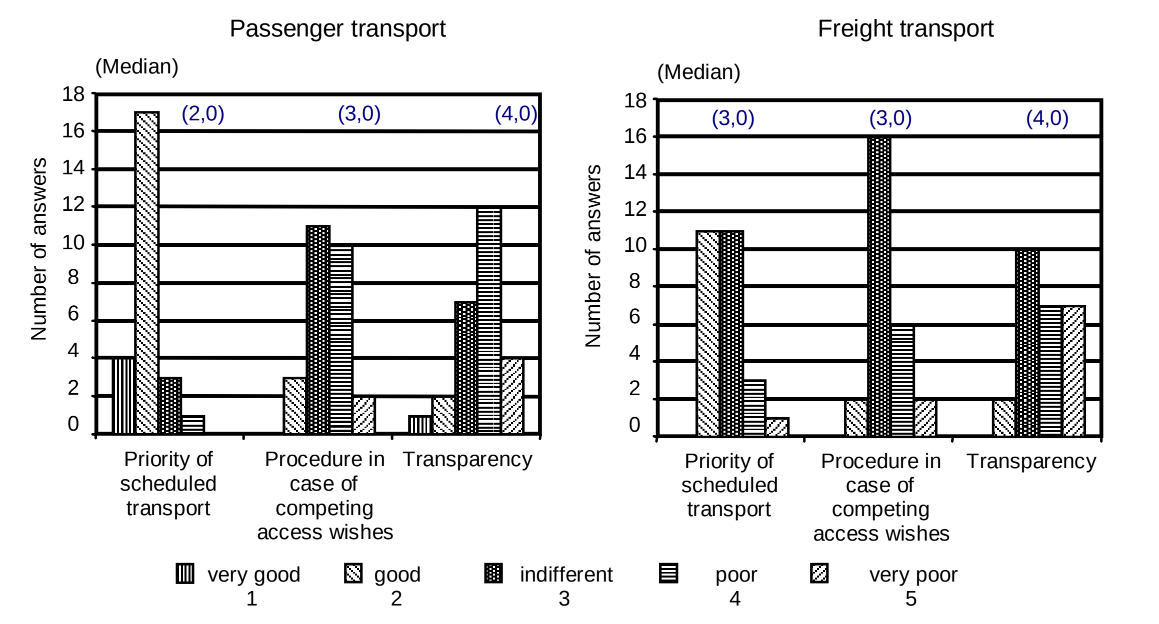


5.3 Track allocation

Half of the DB-competitors have already faced the problem of competing track access wishes. In about 85% of these cases DB Netz has suggested an alternative track access either in terms of time-table or in terms of routes. In most cases theses alternatives were accepted by the competitors. However, in about one quarter of the cases DB Netz refused the track access wish. This concerned mostly competitors in the field of regional passenger transport. There is so far no case known where the bidding process which is foreseen in the network access rules has been introduced.

Concerning the competitors' evaluation of the track allocation figure 5 shows that the general access rules are evaluated as rather indifferent. Surprisingly, even the priority of scheduled passenger transport services in the track allocation procedure is not seen to be problematic even by freight companies which actually suffer by this rule. The most serious problem for the competitors seems to be the transparency of the track allocation process. An interesting result of the survey was also that in contrast to what one could have expected DB-competitors refuse an asymmetric treatment of track access wishes from competitors against those from the DB-companies.

Figure 5
Evaluation of DB's track access allocation procedure and suggestions for reforms



Deutsches Institut für Wirtschaftsforschung

5.4 Information policy of DB to competitors

Missing information is one of the most important problems which DB-competitors face. As stated in chapter 3, DB Netz is not obliged to publish the track access charge system. The company survey showed that about half of the track users are not informed about the route category and the utilisation class of the track they want to use, e.g. they are not informed about the most important price-relevant factors. 12% and almost two thirds of the companies have not been and respectively only partly been informed by DB Netz about the technical parameters of the tracks and the technical requirements of the trains necessary to run at a certain route. Furthermore, information about discounts possible for less track-damaging vehicles, for low-noise vehicles and for time table flexibility was not given to the track users.

The information policy of DB Netz was evaluated as poor by the competitors (see figure 7), a result which was not surprising. The track users expressed the necessity to oblige DB Netz to publish the charge level, all price-relevant factors (route classes, utilisation classes), the possible discounts and also the current state of track utilisation. This latter information is important if competitors want to benefit from the relatively new discount type, the so-called SuperFlex tariff which is granted for using remaining free capacities in the network.

Figure 6
DB's information policy

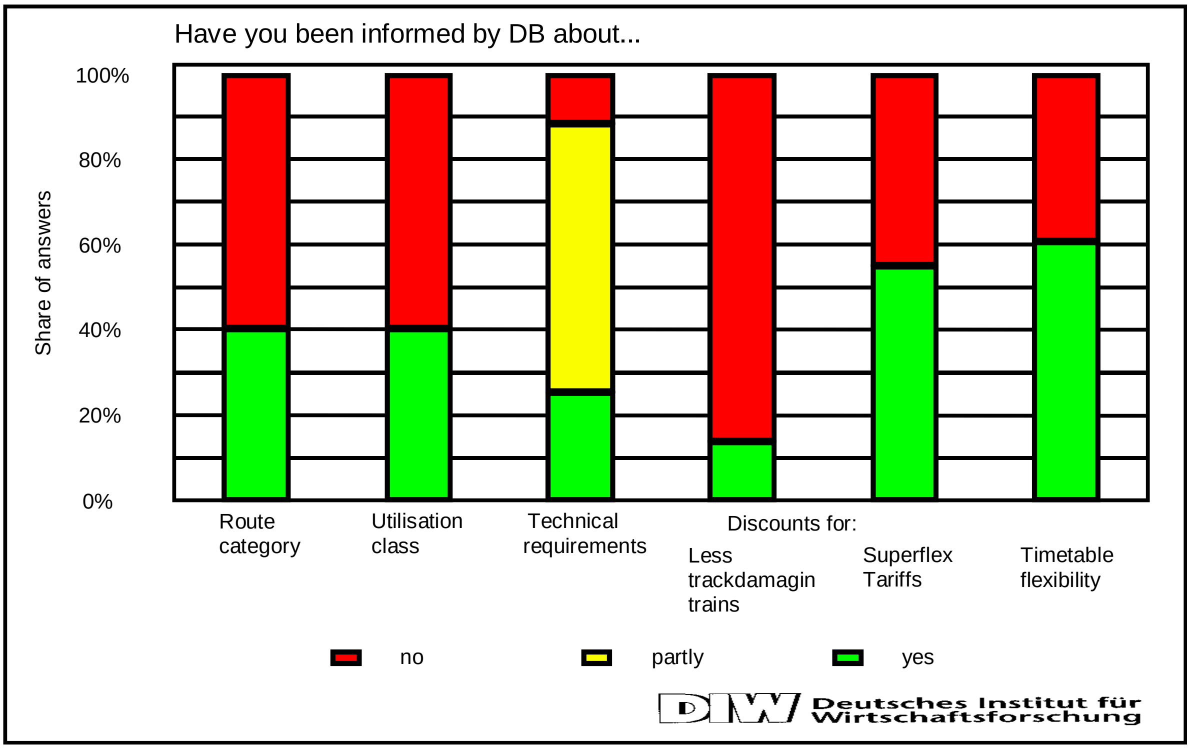
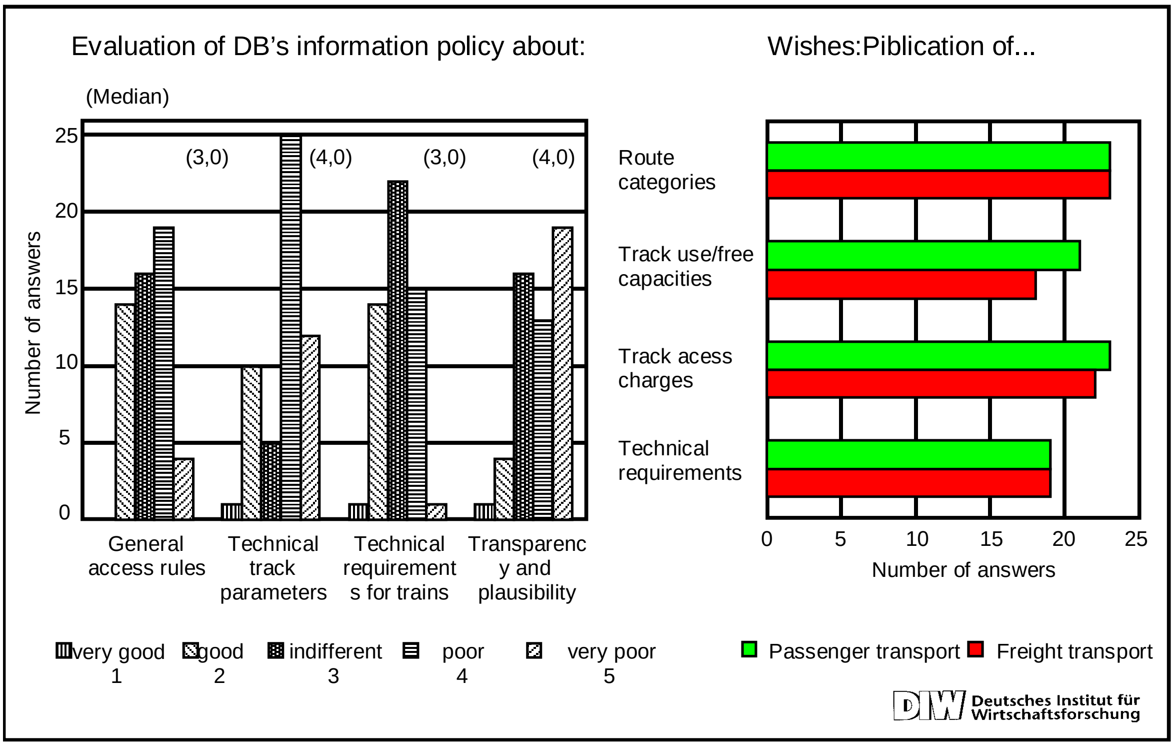


Figure 7
Evaluation of DB's track infromation policy by competitors and suggestions for reforms



5.5 Problems with vehicles and staff

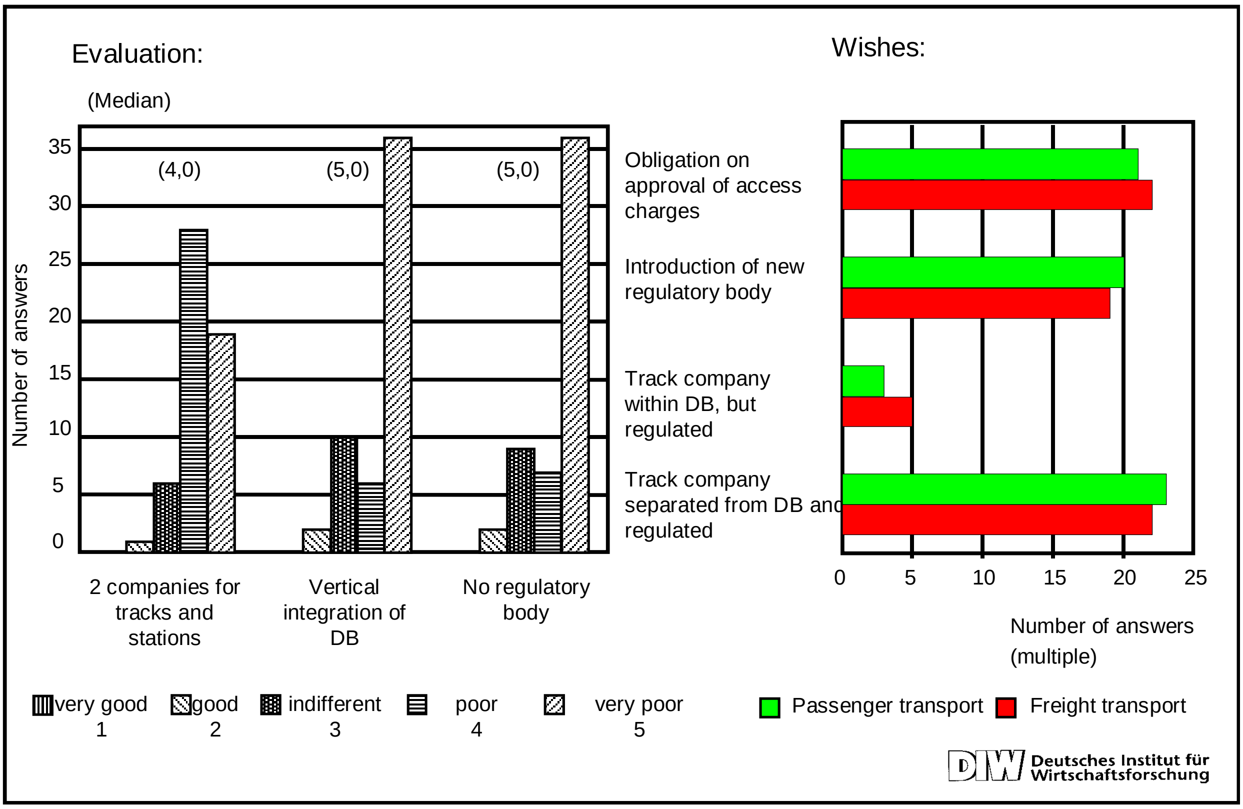
Apart form the problems already mentioned DB-competitors face in particular vehicle problems and difficulties to akquire and qualify their staff. These are mainly problems which do not result from anti-competitive behaviour of DB but rather from general market problems. Locomotive pools for example are just in the process of creation are have just started their business. The only exception of anti-competitive behaviour of DB is the fact that DB sells vehicles only with special contract clauses which do not allow to use the vehicles in competition against DB, or which do not permit for example scrap retailers to sell the vehicles to DB-competitors.

The vehicle problems is particularly severe in freight transport. Two thirds of freight companies evaluated the opportunities to buy vehicles as being poor, the opportunities to rent vehicles were considered to be poor by half of them and the technical parameters of rented vehicles are seen as poor by 20 % of the interviewed freight companies. Additionally, a large share of companies answered to face serious difficulties with respect to staff acquisition and qualification.

5.6 Institutional framework

The track users consider it as extremely problematic that DB Netz operates under one roof with the DB-transport companies. Furthermore, the situation of separate responsibilities for the tracks (DB Netz) and the stations (DB Station & Service) is evaluated to be poor. The track users are additionally in favour of transferring the shunting stations which are currently in the property of their immediate competitor DB Cargo, to the property of an infrastructure provider. A further serious problem is seen in the missing sector-specific regulation of DB Netz. DB-competitors have in most cases contacted the EBA and the Antitrust commission. However, they were not satisfied with these possibilities of an only passive regulation. More than three quarter expressed their request to introduce a new independent regulatory body as a sector-specific one, similar to the one in the telecommunication sector. Furthermore, competitors are arguing for an obligation not only to publish all prices and price-relevant information but also to introduce the obligation of price approval by an independent regulator.

Figure 8
Evaluation of the institutional framework for competition by competitors and suggestions for reforms



6 Conclusions

This paper has presented the findings of an empirical analysis of the market situation in rail transport in Germany with respect to the question to what extent competition on the incumbent's network exists and which role the institutional framework conditions and the track access charging regime play. This paper has identified a considerable discriminatory potential of the incumbent against its competitors which is applied by the incumbent to maintain its dominating market position. This discriminatory potential refers to

- the discriminatory track access charges (DB-favouring design of a non-linear price system),
- the discrimination of DB-competitors by missing information about prices, price-relevant information and technical requirements due to the fact that there is no legal obligation for DB to publish this information,
- the institutional framework, mainly (i) the vertical integration of the infrastructure provider DB Netz with the transport companies of DB under the roof of DB Holding, and (ii) the lack of regulation of DB Netz.

With respect to the current non-linear tariff the conclusion is that it does not fit with on-track competition. Non-linear tariffs for access charges are an economically efficient instrument if budget constraints exist and the 1st best pricing rule of marginal cost pricing fails to cover fixed costs. A further advantage of non-linear tariffs is that they do not require comprehensive knowledge on price elasticities and cross-price elasticities as it would be required by Ramsey-Boiteux prices. In a framework where competition for the market in the form of a time-limited auctioning of a monopoly has been applied (as it is the case in the UK rail market with competitive tendering for franchise contracts of 8 years) non-linear multi part tariffs with self-selection are both a practicable and efficient 2nd best pricing instrument. However, if on–track competition has been introduced non-linear tariffs favour the incumbent and discriminate new entrants. Thus, the decision for the German transport politicians is either to maintain the principle of on-track competition and, consequently, to abolish the non-linear tariff, or to change the on-track competition into a competition for the market and to keep the non-linear tariff. The latter alternative would be suitable for the German rail passenger market. However, it causes problems for rail freight which plays in Germany with a market share of 16% a higher rule than in the UK. Since the freight market is not characterised by scheduled traffic tendering of services seems not to be possible. Thus, if competition in this market is the aim on-track competition is the only solution. Since the rail network in Germany is mainly used by passenger and freight trains different competitional models (on-track competition for freight and competition for the market for passenger transport) and different charging reginmes for freight and passenger transport seem to cause new problems. If thus the on-track competition will be kept as the leading principle in Germany than consequently the non-linear tariff destroys competition due to its discriminatory features. Whether the economic benefits of on-track competition exceed the efficiency losses of not applying a non-linear tariff cannot be answered at the current stage of research and has to remain open.

In any case, independent on the tariff form and even from the competitional model, thwe institutional arrangements have to be changed. In particular the following measures are necessary:

– A genuine institutional separation between the track company and DB AG's transport companies in order to avoid advantageous treatment of the latter in network access and access pricing.
– The introduction of a price regulation for the track company. This requires also to define the type of regulation and the required cost information to be submitted by the track company to the regulator.
– The introduction of a sector-specific regulatory authority with active regulatory power. The responsibilities of the Antitrust-commission, based on the German Act of restraints against competition, would remain unchanged anyway.
– A decision on which type of price discrimination the track company should be allowed with consequences on information to be provided by the track company.
– More transparency of the price system, in particular by making all respective information on network categorisation, price lists, free network capacities and technical requirements public for the competitors.

References

Aberle, G. (1998): "Von der Bahnstrukturreform zum Trassenpreissystem '98", *Internationales Verkehrswesen,* Vol. 50, No. 10.

Armstrong, M. *et. al* (1996): "The Access Pricing Problem: A Synthesis." *The Journal of Industrial Economics,* Vol. XLIV, No. 2.

Baumol, W. (1983): "Some Subtle Issues in Railroad Regulation." *International Journal of Transport Economics* 10: 341-355.

Boiteux, M. (1971): "On the Management of Public Monopolies Subject to Budgetary Constraints", *Journal of Economic Theory,* 3.

Braeutigam, R. (1980): An Analysis of Fully Distributed Cost Pricing in Regulated Industries. *The Bell Journal of Economics.* Vol. 11, No. 1.

Faulhaber, G. R. and Panzar, J. C. (1976): "Optimal two-part tariff systems with self-selection." *Bell Laboratories Discussion Papers,* Nov. 1976.

Feldstein, M.S. (1972): "Equity and efficiency in public sector pricing: The optimal two-part tariff." *Quarterly Journal of Economics,* Vol. 86 (1972), pp. 175-187.

Haase, D. (1998): "Das neue Trassenpreissystem der Deutschen Bahn AG." *Internationales Verkehrswesen,* Vol. 50, No. 10.

Knieps, G. (1998) "Das neue Trassenpreissystem: Vorteile eines zweistufigen Systems." *Internationales Verkehrswesen,* Vol. 50, No. 10.

Laffont, J.-J. and Tirole, J. (1994): "Access Pricing and Competition." *European Economic Review* 38: 1673-1710.

Link, H. (1994): “Structural Reform of Germany’s Railways - Could Japan Serve as a Model?”, *Economic Bulletin of DIW*, Vol. 31, No. 11. Berlin.

Link, H. (1997):”Access Pricing in the German Railway System - Are the Track Charges Cost-covering?” Paper presented at the Fifth International Conference on Competition and Ownership in Land Passenger Transport. Conference Proceedings. University of Leeds.

Link, H. (1998): “The German railway market in transition - Is there a need for regulation?”, Paper no. 896 Proceedings of the 8th World Conference on Transport research, Antwerp 1998.

Ramsey, F. (1927): A Contribution to the Theory of Taxation. *Economic Journal,* Vol. 37.

Spence, A. M. (1977): “Nonlinear prices and welfare”. *Journal of Public Economics*, Vol. 8, No. 1.

Willig, R. (1978): “Pareto-superior nonlinear outlay schedules.” *The Bell Journal of Economics*, Vol. 9, No.1.

Willig, R. (1979): “The Theory of Network Access Pricing”, In: H.M. Trebing, ed., Issues in Public Utility Regulation. Michigan State University Public Utilities Papers.

Short Term Adjustments in Rail Activity

a Case for Central Programming

Emile QUINET

CERAS-ENPC, 28 rue des Saints Pères, 75007 Paris, France.
tel. : 00 33 1 44 58 28 74
fax : 00 33 1 44 58 28 80
email : quinet@mail.enpc.fr

Abstract

In many countries and especially in Europe, Railways reforms have fragmented the rail industries between one Infrastructure Manager and several Rail Operators, and raised new interest on Railways short term adjustment, a procedure hitherto run inside the integrated historical operator. A current stream of thought often bases it on the same tool as the tool recommended for road short term adjustment, namely infrastructure charging. The argument is that, due to the differences between road and rail markets, pricing cannot be a proper way to achieve an optimal allocation of rail resources ; centralized programmation, taking into account various externalities, is the right mean for that objective ; information asymetry on private values of the services for Rail Operators could be coped with through auctions, but these auctions should be understood as a tool for central optimization and cannot not replace it. These considerations leads to assess the right role of infrastructure charges in rail, and, taking into account the present know-how, to draw some conclusions about research fields to explore and means to improve the present practices.

Keywords

Rail fragmentation ; short term adjustment ; infrastructure charges ; path allocation ; decentralization ; centralized programming.

1. Introduction

A lot of research have been done about railways reforms. The relations between the Infrastructure Manager and Railways Operators have been especially studied in Europe where the reforms launched by the European Union are based on the unbundling between infrastructure and operations. Relations between these two types of actors are manyfold, and the following thoughts will focus on short term adjustments : pricing and path allocation.

The bulk of the arguments has focused on infrastructure pricing, following the case of road, where pricing allows for the optimization of the use ; the core of the doctrine is that Short Run Marginal Social Cost (SRMSC) should be the basis for infrastructure pricing.

The main idea underlying this paper is that rail activity is quite different from road traffic ; the differences lie into the nature of congestion, the extent of social values compared to private values, and the oligopsonistic nature of the market for infrastructure use ; the problem of short term adjustment in rail is akin, not to the road short term adjustment, but to the problem of infrastructure investments programming in a network.

The argument will be developped along the following steps : the following section (section 2) reviews the current European practices and trends in railways short term adjustment procedures and shows how they are inspired by road experience. The purpose of the third section is to show that the rail short term adjustment problem is more similar to the problem of investment infrastructure choice than to the problem of road short term adjustment. The fourth section explores the role of prices. The fifth section shows that auctions can alleviate the task of private value revelation, though they are made difficult by the complexity of the constraints which they have to comply with. The sixth section explores the presently used solution : priority rules ; it is shown that this solution is perhaps the best operational solution for the present but should be refined. The conclusive section deduces from the previous developments the implications for research and policy

2. The current European practices for short term adjustments.

European practices for short term adjustment have been modified by the railways reform launched by the European Commission.

The basis of this reform is to create competition in railways ; it is aknowledged that infrastructure management is an increasing return to scale activity and cannot be opened to competition ; so the core of the reform is to undbundle infrastructure and operations and to set up competition in the operations[1]. The reform have taken various shapes according to the countries. For instance in UK the Infrastructure Manager (IM) is a private company the activities of which are controlled by a regulator, and Rail Operators (RO) are private firms, franchized for passenger traffic ; there is no on-track competition on the passenger market, just competition for the market ; the on-track competition on freight market is very low, one firm has almost the whole market share. In Germany the IM is a branch of the hitherto public holding which holds also the Rail Operators (RO) issued from the historical operator DB ; there is on-track competition for freight and long distance passenger traffic and competition for franchize in local passenger traffic.

The reform has been different in other European countries, but everywhere the IM has to procure infrastructure to the ROs through market relations. This procurement has two sides : on the long run the problem is to invest on (or close) new infrastructures ; on the short run, the problem is to price and allocate the paths to the various ROs.

The goals to reach through these procedures are (Nash 2000) :

- To increase rail market share, especially *vis-à-vis* road
- To avoid discrimination between ROs and especially between the incumbent (which often maintains some links with the IM) and the new entrants.
- To foster competition in the operations sector.
- To induce the best use of infrastructure

[1]The 91-440 Directive imposes competition just for international combined transport and transit freight transport. But many countries have liberalized rail industry to a much higher degree

- To raise funds for infrastructure maintenance and development
- To achieve public goals such as public service obligations or distributional effects.

Infrastructure charging is the main tool on which attention has been drawn. This point is quite clear when looking at the amount of papers devoted to pricing *vis-à-vis* path allocation : pricing has been dealt with in the Green Paper (European Commission 1995), the White Paper (European Commission 1996) ; these texts have set up recommendations which are direct extensions of the standard economic theory of road short term adjustment : the short run marginal social cost (SRMSC, including : infrastructure marginal cost+congestion cost+environmental and safety costs), insures first the clearing of the market and second the optimal resources allocation ; SRMSC can possibly be amended in order to allow for budget balance constraint or to cope with market imperfections ; Ramsey-Boiteux pricing or two-parts tariffs are the main solutions to these situations.

Each country has issued regulations about infrastructure pricing. These tariffs (NERA 1998) generally takes into account :

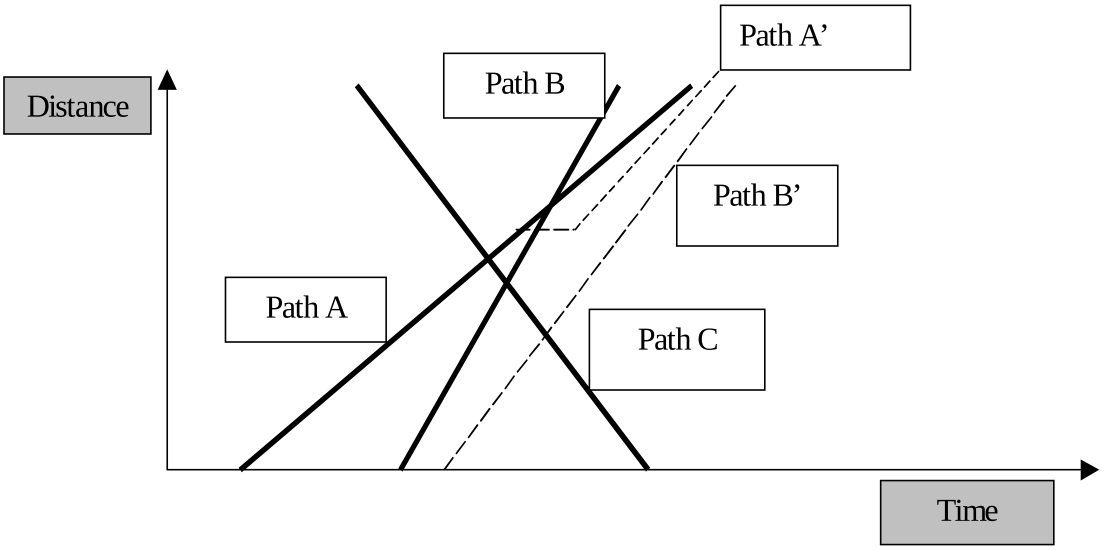
- A fixed term per km
- A term related to the size of the train (tons, number of passengers) and/or to the type of the train (passenger, freight), to the type of track (HST, other) and to the time (day, night..)
- A reservation term, depending on the same parameters as the previous one

These principles have taken various forms in each country : In UK for instance, the infrastructure pricing is composed of a lump-sum negotiated term with each RO for the base service and a small marginal term for each additional train ; a performance regime is set up, implying bonuses and maluses for delays and on-time services to be payed between the ROs and the IM. In Germany, there is a menu of two tariffs, between which each RO has to choose : a proportional one, fit for small ROs, and a two part tariff, fit for big ones. In France, the infrastructure charges cover about 50% of maintenance costs and 30% of total expenses though in UK they cover 100% as well as in Germany.

This lot of attention devoted to infrastructure charges is probably due to the fact that it is a new problem, hidden up to now by the bundling of IM and RO, and, given the road example, it was thought that its solution would solve the whole problem of short term adjustment in railways ; in fact, it is not now the case, and there are still applications for uncompatible paths from different ROs (see Box 1). This drawback is generally thought as being due to our poor knowledge of SRMSC, especially congestion costs.

In comparison, little attention has been payed to path allocation issues, probably because it is not a new problem : it already occurred inside the historical incumbents. The European Union has just mentioned it in some directives, and the presently discussed Directive (European Commission 1998), which aims at dealing this question, just fixes the procedures to be applied : no discrimination between the appliants, especially not in favour of the historical incumbent ; a precise time-table for the applications ; the allocation should be made by an independant body. The algorithm for solving conflicts is not precised ; in fact all countries have established priority rules which are in the line of the procedures used by the historical incumbents.

Box 1
Incompatibilities between paths
(supply externalities)



This graph represents several paths a one-track line between stations X and Y. Applications have been made by three ROs for services A, B, and C. A and B are going from X to Y and C from Y to X. These three paths ar e incompatible.
The incompatibility between A and B can be removed either by delaying service A at the middle of the link (path A') or by shifting path B (path B').
The incompatibility between paths A and B and path C is very high ; it may happen that in such case, delaying or shifting would rreduce too much the value of the path, and in that case there is no other solution than to cancel either path C or paths A and B.

These priority rules generally follow the hierarchical order :

- International and intercity passengers
- Local passengers
- Goods trains, among which : first combined transport, second other goods trains.

They are implemented through a large use of « grand-father rights » and incremental changes from year to year ; these priority rules do not include path reservations for infrastructure maintenance purpose, these ones being decided by the IM according to historical tradition. There are of course country specificities : In UK and Danemark, some subsidies can be granted to combined trains when they cancel environmental damages from the road.; there is a possibility of several years reservation with confirmation every year. In Germany and Switzerland, within equal priority rank for conflicting applications, the one which proposes the higher payment is choosen. In Sweden, priority rules take into account the social value of the service. In all countries path allocation is made by the infrastructure manager, under the control of the rail regulator.

3. The rail short term adjustment is more akin to investment programming than to road short run optimization.

The emphasis given to infrastructure pricing versus path allocation is probably influenced by the road case where it is proved that infrastructure pricing ensures the optimal allocation of resources. But there are large differences between road and rail short term adjustments, related to the difference of the market structures :

In the classical presentation, road demand is a final demand market, characterized by a continuum of agents all independants with possibility of quick entry or exit from the market. In such a situation, prices have very interesting and specific properties which justify their use as the optimal and unique tool for short term adjustment issues : there is a price, the same for every user (no discrimination) which induces an optimal situation from a collective point of view. This optimal price can be attained through a learning or trial-and-error process lead by the regulator on an incremental basis. This price eases the information exchanges which are necessary to solve the allocation process and allows for a partial decentralization of this process : instead of telling every user what he has to do, the planner has just to announce a price and each user takes his own decision after this announcement ; then the planner checks if, taking into account the decisions of the users, the market clears properly ; if it is not the case, he makes another announcement higher or lower according to the unbalance of the decisions of the users ; this procedure induces huge gains in transaction costs compared to the centralized planning procedure. Of course, the regulator has to know a private information parameter (the values of time of the users) to calculate the optimum ; but, in road traffic, this parameter can be known by means of inquiries without strategic bias from the users.

Situation in railways industry is quite different : there are few operators, the market being an oligopsone, the operators expressing an intermediate demand, and there are a lot of externalities and interactions inside the market , both on the demand and on the supply sides.

On the supply side, rail congestion is a much more complex phenomenon than road congestion ; in this mode, the congestion is related, at least in the classical analysis, just to the number of users ; in rail the congestion is related, not only to the number of users, but also to the speeds of the users (see Box 1).

On the demand side, many services are either substitute (for instance the services proposed by competiting ROs) or complements (for instance positive externalities between a trunk service and its feeders). Besides, in road, demand WTP can be expressed through rather simple parameters (for instance the demand elasticity or the value of time distribution), but in the rail industry, ROs'WTP (which is their profit) is not easy to reckon and its revelation is subject to strategic behavior, a problem which does not appears in the road transport. Finally, in road transport, the demand WTP contains the whole surplus of the demand side, which is not the case in rail transport ; in this last case the demand surplus also comprises the final demand surplus.

The objectives of short term adjustment are the same in both modes : to elicit the solution (volume of traffic in road, path allocation in rail) which bears the higher collective surplus. But the tools have no reason to be the same.

Letting apart the problem of the gap between collective surplus and WTP of the demand, let us have a look at the virtues of prices in eliciting the optimal solution. The examples contained in box 2 show that :

- Due to the discrete feature of the problem, prices eliciting a solution are multiple.
- It may happen that no simple price for each block clears the market properly, and that it is necessary to set prices for combinations of blocks which are not the sum of the prices of each elementary path composing the whole path.
- So it is necessary to know the optimum in order to fix the prices ; prices cannot alleviate the exchange of information in the optimization process and they cannot help for the solution of path allocation.

In fact choosing a path allocation is similar to programming a set of infrastructure investments in a network. Of course the size of the problem is not the same ; in investment choices, the decision at stake is related to all the services running the infrastructure ; in path allocation the decision at stake is related to a single service ; besides, there is no use of dicounting in the path allocation problem and their is no irreversibility.

But similarities are numerous. In both situations, the items to assess are services or clusters of services. It is clear that it would not be sensible to device the infrastructures with a criterium based on services value, and to device the services it provides with another criterium when the infrastructure is built ; the criteria should be the same, i.e. the collective surplus, composed of the IM surplus, the RO surplus, the final demand surplus, and possible other external effects[2]. As there is no a-priori restriction on the complementarity-substitution relations between the services, there is no simple optimization algorithm for solving this problem of optimization. In the most general case, there is no other procedure than to enumerate the possible solutions and to choose the one which yields the maximal collective surplus. In particular, incremental procedures may lead to sub-optimum, as it is shown in the box 3.

This statement raises two issues. The first one is related to infrastructure charging : what is its role, how should it be calculated? The second one is related to the information requirements. it is necessary to determine the collective values of the possible services. These two issues will now be successively adressed.

[2]In this framework, it is tempting to define “the collective value” of a service as the difference between the NPV of the program with the service and the NPV of the same program without this service. But it is also clear that this “collective value” may be different according to the base program, except in the case when the service is independant of all other services for any program ; the unambiguous definition of thecollective value would need to precise that the base program is the optimal one. For sake of simplicity, we will nevertheless use this expression in the rest of the text.

Box 2
Price decentralization issues

- <u>Multiplicity of prices eliciting the optimal solution</u>

Assume that the same path is subject to applications from two operators A and B, which value it respectively 10 and 20. Assume also that there is no other conflict concerning this path and that the values of this path are independant from the rest of the time-table. Then any price between 10 and 20 would decentralize the optimum.

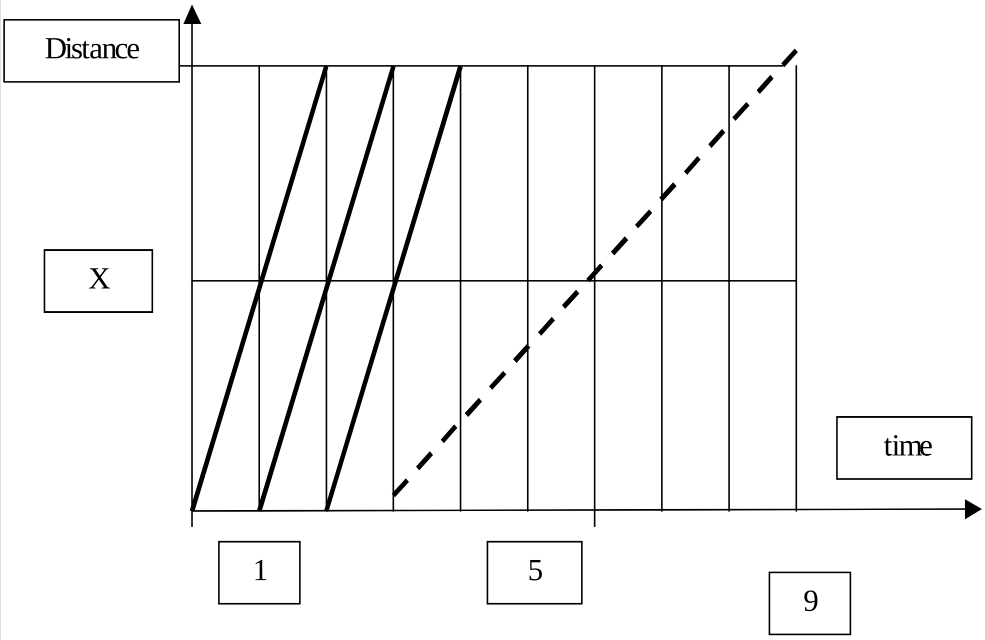
- <u>It may happen that no simple price per block clear the market.</u>

Let us assume (Caillaud 2000) that two operators A and B are applying for two neighbour blocks 1 and 2, these two blocks being related to the same link but differing by the time. A is an operator of slow trains, cannot use a single block and needs to have the two blocks ; for it, one block (either 1 or 2) is valued 0 and the two block together are valued 600. For B, which is an operator of fast trains, can run a train with one block but has no use of two adjacent blocks : two close services would not get more patronage than one block ; so, for B one block (either 1 or 2) is valued 400, and the two blocks together are valued 500. It is clear that the optimal allocation is to give the two blocks to A. But no simple price system, P1 for 1 and P2 for 2, implement it ; these prices should be such as :
P1>400
P2>400
P1+P2<600
which is impossible
In order to select the optimal allocation through prices, it would be necessary to set the following prices :

- for one single block, a price higher than 500
- for two blocks taken together, a price between 500 and 600

It is clear that in order to set up this result, it is necessary to have first solved the optimization problem ; after this first step, the allocation of path can be achieved through autority, and does not need to use the price procedure.

Box 3
Sub-optimality of incremental process



Let us assume that the space-time is composed of 2*9 elementary paths (blocks), as shown in the graph. The compatibility constraint states that a block should not contain more than one path.
Two types of paths are in conflict : Path A and Path B. These paths have independant values , respectively 8 and 30. The possible allocations are :

- (a) : 2 paths B. Value : 60
- (b) : 1Path B and 3 paths A (the allocation which is shown on the figure). Value : 54
- (c) : 0 path B and 8 paths A. Value : 64

It is clear that the best allocation is (c). Nevertheless, if the planner starts from the non-saturated situation where there is just one path B, and sticks both to the grand-father rights and to the incremental logic, the result is the sub-optimal allocation (a). Furthermore, incremental change from (b) leads to (a) and do not move towards the optimal situation (c).

4. The role of infrastructure charging.

What is then the role of price in rail short term adjustment? In order to understand this point, it is good to go back to the analogy of infrastructure investment optimization. Let us take for instance the case of a narrow valley where it is possible to build either a road or a railway track, but not both of them, a situation analogeous to the situation of conflict in path allocation. The choice between them has to be done through a cost-benefit analysis ; when the choice is done, let say in favour of the road, the pricing of the road use does not depend on the foregone alternative, here the railtrack, it will depend just on the situation of the road. Similarly, the choice between two incompatible rail services has to be made through a cost-benefit analysis, and when it has been achieved, the pricing of the choiced service does not depend on the characteristics of the foregone service.

It has to ensure fairness in competition between modes and to ensure a good allocation of infrastructure maintenance, not to choose between two services. So it has to take into account : infrastructure costs, second best issues: budget constraint, non-optimalities in related markets, other externalities such as environmental damages, and possible distributional issues. It may be noticed that in this case, no congestion cost would be integrated to the railways pricing. Does it mean that these congestion costs have no interest? Certainly not ; in fact they play an importance part in the calculation of the social values of the services in the framework of the cost-benefit analysis, but after this step they play no role in the following step of pricing.

In the previous analysis, infrastructure has been considered only as a possible mean to achieve efficiency ; charging contribute to other objectives, namely to achieve equity considerations and to raise funds for the infrastructure manager. These two objectives can be achieved through the classical tools, and especially for the second one, through RAMSEY-BOITEUX pricing or price discrimination. Funds can also be raised through a device which helps to solve the problem of revelation of private values : auctions.

5. Calculation of collective values and revelation of private values.

This planning process is hampered by the imperfect knowledge of the collective values to take into account. The collective values cover both consumers'surplus, external effects and the profits of the ROs. None of these three elements can be neglected, as the experience of investment appraisal shows it : it is quite current that the consumers'surplus and external effects amount to about the half of the total value of investments. How to calculate each of them, and especially how to avoid strategic bias in the private value?

Let us first consider the profits of the ROs ; the asymetry information between the planner and the ROs induces to set up a revelation process. Auctions have been studied in this objective. The main research in this field have been done by NILSSON (1996 and 2000) and by BREWER and PLOTT (1996). The corresponding studies have taken the form of experiments ; the subjects are given values for some paths and they participate to auctions ; in these experiments, auctions do not prevent from centralized planning of the optimal services but help to reveal the private values (profits) of the ROs through an iterative process :

- Bidders (the ROs) are asked to announce values for the paths they want. Building on these values, the planner optimizes the total surplus and publishes the result of the allocation.

- Knowing this result, the ROs make other bids which are used to reckon a new optimum allocation, the results of which are published.

- And so on till no RO wants to change its previous bid.

The experimental studies have used different sort of auctions (first or second price..). They have been achieved on simple networks with no demand externalities, just supply incompatibilies ; in these cases the bids are rather simple : the bidders have just to bid on the services they want, and to tell how their bids are modified if the services are moved forward or backward ; the results of the experiments are close to the optimum (which can easily be calculated, as the planner has the true private values of the bidders).

These results are encouraging, but they must be improved in taking into account demand externalities, for instance in the case of two ROs operating competing services. In that case, the bidders have to make bids for the paths they want to get, conditional on what is obtained by the other bidders for the other possible paths. It is clear that in that situation the bids are much more complicated, it is a question to know whether such auctions are workable. Assuming that it is the case, auctions would solve the problem of private values revelation[3].

Social value of the paths includes also environmental external effects (noise, pollution, safety..) and consumers surplus. Environmental external effects can be coped with either through subsidies (or taxes) applied to the services, or through negative or positive handicaps added to the bidded services for the optimization procedure achieved by the planner.

The estimation of consumers'surplus is more complicated. It is not clear whether the ROs have or not a better knowledge than the planner ; if we consider that the bulk of this consumers'surplus is made of time savings (in the case of passenger traffic), the value of time is probably a common knowledge ; the estimation of the number of hours saved is perhaps better known by the ROs, but the information asymetry should be assessed and is perhaps not very large. Another point is to check whether the consumers'surplus is or not well correlated with the private value of the RO (profit). If the answer was that the correlation with profit is high, then the evaluation of the consumers'surplus could be done along with the auction procedure ; in the present state of the art, it seems that this question stays open.

As a whole, it appears that auctions do not suppress the need for centralized planning, but can cope with the problem of private information on ROs'profits, under the provision of the results of future research,and that the other parts of the social value of the services contain probably less information asymetry, this point being also subject to further research.

6. Priority rules should be refined.

As long as the results of these research are not known, the actual procedures, which are based on priority rules, can be a good transitory compromise. Of course, they do not solve the information asymetry on private values ; but they can take into account the social value, as

---

[3] Let us note that the IM should participate to the auctions as it has to use some paths for maintenance purpose, and trade-offs have frequently to be made between an organization of maintenance which is cheap but use congested paths, and another one which is more expensive but takes place in non-congested paths.

calculated by the planner, under the form of handicaps in optimization determination. They present the advantage that transaction costs are low.

Presently they are rather rough, but can be refined. They can be calculated through indexes, representative of the social value of the service, classified for instance according parameters like the following ones :

- The type of service
- The level of traffic of the service
- The conditions of the competitive mode
- The flexibility of the time-table

…etc

These indexes could be injected into more or less sophisticated optimization procedures, such as the solution of binary local conflicts in a first step, or complete optimization programs in an ideal situation.

7. Possible extensions, research and policy implications.

The previous considerations have shown that short term adjustment in railways is not similar to the road one, due to the differences in the structure of the demand. This short term adjustment is, from a logical point of view, composed of two steps. One is path allocation ; this problem cannot be decentralized through prices, it is quite similar to infrastructure investment programming in a complex network, and must be solved through centralized planning taking into account the collective surplus of the services, namely ROs profits, externalities and consumers'surplus, these last two parts being not at all neglectable ; the information asymetry on ROs profits can be alleviated through auctions, which are especially difficult to implement when there are demand externalities, i.e. complementarities or substitutions between services.

The infrastructure pricing is the other step ; infrastructure charges should include infrastructure marginal cost, external environmental cost, possible corrections of market imperfections, and devices to cope with budget constraints or cost of public funds (for instance Ramsey-Boiteux pricing), but no congestion cost.

The revenue of the IM would be composed of the revenues from the infrastructure charges and the revenues from the auctions[4].

In the present state of knowledge, we are not able to seize the magnitude of the real information asymetry in collective surplus estimation, the nature, size and effects of supply incompatibilities and demand externalities. So it is wise, on the policy field, to use

[4] It must be noted that these conclusions are not limited to the case of railways, but can be applied to many other markets. In fact, they are valid whenever the demand market is an oligopsone with no possible entry and when there are both supply incompatibilities and demand externalities, i.e. when the utility of each user depends, not only on his/her allocation, but also on the allocations of the other demanders (slot allocation for air transport, for instance).

simultaneously all tools and, on the research field, to explore the real scope of the possible theoretical particularities of the problem. More precisely :

- To develop research upon auctions, and achieve partial implementations of auctions.
- To develop the calculation of “collective values” of the paths, and especially our knowledge of the effects of introducing a new path in the time-table, i.e. the congestion costs.
- To explore the restrictions which can be imposed on the demand externalities, the possible solutions to the optimization problem departing from incremental changes, and to input the result into the calculation of the social value (iterative process).
- To refine the priority rules in order to take better account of “social values” of the paths.
- As long as the auction procedure are not operational, to adjust the charging system in order to take into account an approximation of congestion and scarcity costs in infrastructure costs.

The previous reflexions leave apart many problems such as how to induce competition and entry on the market (possibly long term contracts in order to foster commitment from possible entrants, but not too long in order to adapt to possible changes, and with penalties for non-use of the reserved path in order to prevent from predative behavior), or the institutions best fit to run such process (the fact that IM should take part to the auction process and the importance of social versus private values advocate for running path allocation through an independant public agency).

Aknowledgements

The ideas developed here have been induced by the work of a team animated by the author, composed of M. Carrier, J.N. Chapulut, C. Charmeil and C. de Fenoyl; this team has been asked by the Ministry of Transport to make proposals to improve path allocation procedures. The author is indebted to commentaries on a preliminary version made by participants of the seminars : “Railroad Conference” organized by IDEI (Toulouse) and CERAS (Paris) and held in Paris-La Défense, July 2000 ; and “Rail Infrastructure Charging” organized by VATT and ITS and held in Helsinki, July 2000. As for eventual mistakes or misinterpretations, the usual disclaimers apply. Moreover, the views expressed are own and do not represent in any way the views of the institutions the author is affiliated with.

References

Brewer, P. J. and Plott, C. R. (1996). A binary conflict ascending Price (BICAP) Mechanism for the decentralized Allocation of the Right to use Railroad Tracks. *International Journal of Industrial Organization*, Vol 14, N° 6, pp857-886.

Caillaud, B. (1999). *Allocation des sillons ferroviaires : problématique économique*. Internal report, CGPC-CERAS, Paris.

European Commission (1995). *Green Paper. Towards fair and efficient Pricing in Transport*. COM (95) 691.

European Commission (1996). *White Paper. A strategy for revitalizing the Community’s Railways*. COM (96) 421.

European Commission (1998). *Proposal for a Council Directive relating to the allocation of railways infrastructure capacity and the levying of charges for the use of railway infrastructure and safety certification.* COM (98) 480.

Nash, C. (2000). *Recent developments in Railways Reform in Europe*. Railroad Conference, Paris, June 2000.

NERA (1998). *An Examination of Rail Infrastructure Charges*. Report for the European Commission, DG VII, London.

Nilsson, J. E. (1996). Allocation of track capacity. Experimental Evidence on the Use of Priority Auctioning in the Railway Industry. Forthcoming in *International Journal of Industrial Organization.*

Nilsson, J. E. (2000). *Towards a Welfare enhancing process to manage Railway infrastructure access.* Railroad Conference, Paris, June 2000.

ESTIMATION OF MARGINAL INFRASTRUCTURE COSTS OF THE FINNISH RAIL NETWORK

Tiina Idström, Finnish Rail Administration

Abstract

The purpose of this study is to estimate marginal costs of railway infrastructure of the Finnish rail network. The paper contains only a brief summary of results; the more detailed report will be published by the Finnish Rail Administration (RHK). This report will also be a graduate work for the University of Jyväskylä.

The approach in this study is similar to the Swedish study "An Economic Analysis of Track Maintenance Costs"[1] (Johansson, Nilsson 1998). The Swedish study had maintenance costs as an independent variable, in this report also renewal costs have been analysed, like the European Commission High Level Group has recommended. The analysed gross section data consist three years 1997, 1998 and 1999. The results of analyses provided indicators of scale economics in track use. Furthermore, average marginal cost estimates had been derived. As expected, these estimates, if used as marginal cost pricing tool, would lead quite low cost recovery levels. The results were quite same as the Swedish ones. Another conclusion was that the used model could explain better variance of maintenance cost than variance of replacement investment costs.

1 Introduction

The European Commission White Paper "Fair Payment for Infrastructure Use"[2] introduced a phased approach to a common transport infrastructure charging framework in the EU. The paper proposes to apply the principle of Social Marginal Cost Pricing. According this White Paper transport infrastructure charges should strictly be based on marginal cost of infrastructure, congestion, accidents and environmental damage.

But how Social Marginal Costs should be specified in practice? For the each social cost category the has been further considerations. The expert advisor's recommendation of infrastructure cost definition are presented in report "Calculating Transport

---

[1] Johansson P. ja Nilsson J-E.(1998): An Economic Analysis of Track Maintenance Costs, Economics Department, School of Transport and Society, Dalarna University, Paper no 881 for the 8th WCTR

[2] European Commission (1988): Fair Payment for Infrastructure Use: A Phased Approach to a Common transport Infrastructure Charging Framework in the EU, White Paper, Directorate- General VII Transport, Directorate B- Inland transport

Infrastructure Costs“[3] (here on this report is referred as a HGL-report). The report develops a methodological framework for the estimation of marginal infrastructure costs focusing on road and rail transport.

The purpose of this study is to estimate marginal infrastructure costs of Finnish railway. The study seeks to apply HGL-report recommendations of cost estimation. In addition the used approach is similar to Swedish study “An Economic Analysis of Track Maintenance Costs“. This similarity enables to compare results from the Finnish data to the Swedish results.

The study begins with the description of analysed data. Firstly, the relevant cost for the marginal cost pricing are defined by the recommendations of the HLG- report. Secondly, explanatory variables are summarised. Section 3 contains the model description and section 4 provides estimation and associated tests. The results of the study are reported in section 5 and concluding comments are presented in section 6.

2 Data

2.1 The analysed cost data

The HLG-report contains recommendations how transport infrastructure costs should be calculated. From the efficiency point of view only variable cost i.e. costs, which are depending on traffic volumes, are relevant for pricing. Therefore, one key issue in report is to elaborate a transparent distinction between variable and fixed infrastructure costs. The proposition of report recommends that land purchase, construction of new lines, upgrading/enlargement of existing lines and administration’s overhead should be regarded as fixed cost and thereby not to be included into calculations. Whereas other cost categories, like renewal and construction maintenance, are defined to be at least partly variable.

The HLG-report also recommends that railway stations and freight terminals would be excluded from the analyses. The reason is that these types of infrastructure are characterised by other kind of cost behaviour/cost function than “transport ways“. Therefore if the use of rail stations and freight terminals is to charged another methodological frameworks have to be developed.

In this study the recommendations of HLG-report are interpreted in a way, that costs which are classifies to be fixed are excluded from the calculations. Whereas the cost which categorised at least partly variable costs are analysed. Also the cost of railway stations and freight terminals are left out. This is due to recommendation, but also the fact that there is not available gross tons data for stations and freight terminals. Table 1 summarises excluded costs. Majority of fixed cost are caused by upgrading/ enlargement of existing lines. These costs were 209 (1997), 337 (1998) and 508 (1999)

[3] European Commission (1999): Calculating Transport Infrastructure Costs, Final Report of the Expert Advisors to the High Level Group on Infrastructure Charging (Working group 1)

million FIM. There is no track section data of the traffic control costs; therefore also these cost were forced to leave out.

Table 1: Fixed costs and other not analysed costs (million FIM)

Cost category	1 997	1 998	1 999
Fixed cost	408	598	782
Stations and freight terminals	220	210	180
Traffic control	200	210	207
The sum of not analysed costs	828	1 018	1 169

In turn, table 2 contains the analysed cost data. Values in the row “Maintenance costs“ are used, when calculating traffic dependence of Finnish maintenance costs. These cost are symbolised with C_1. The other analysed cost variable is C_2. This variable is sum of maintenance, minor repairs and replacement investments costs. HLG-report classifies that all these cost should be taken account, when defining marginal cost of railway infrastructure. These cost are called as HLG-costs.

Table 2: The analysed cost data (million FIM)

Cost category	1 997	1 998	1 999
Maintenance costs (C_1)	277	251	252
Minor repairs	112	102	105
Replacement investments	930	1 033	840
The sum of analysed costs (C_2)	1 319	1 386	1 197

According to the financial statement information the total costs of Finnish Rail Administration (RHK) were 1898 (1997), 2049 (1998) and 2115 (1999) million FIM. The HLG-cost category C_2 presents 69.5% (1997) 67.6% (1998) and 56.6% total costs of the financial statement. In reality percentages are slightly higher due to accounting differences in the financial statement data and the data used in this study.

2.2 Track unit separation

The data used in this study is from RHK´s accounting system, VR's accounting system and statistics (VR is Finnish train operator). Each data source has a different way to separate the rail network into track sections. For the regressions the network has to be divided to homogenous track sections. Here homogenous track sections are called track units. The basic principle for track unit modification is that track units should have homogenous traffic volumes and technical quality. In this study track unit specification is based on gross ton data's track separation and maintenance level information. According to these data sources the Finnish rail network is divided to 93 track units.

2.3 Characterisation of variables

Table 3 gives description of variables used in this study. It is notable that there are two different cost variable in the table, namely maintenance costs C_1 and HLG- costs C_2. These cost variables are used separately as a dependent variable. The data regarding maintenance costs is based on VR's accounting system. VR reformed its accounting system in 1997 and therefore earlier data is incomparable to present one. Thereby, the analysed data consists only years 1997-1999.

The HLG-costs parameter contains also replacement investment and minor repairs costs. The data of these costs is from RHK's accounting system. In RHK's accounting system cost are registered by orders not by track section. However, most of the orders contain also information, which part of rail networks these orders are to be done. Majorities of the order expenditures are caused by replacement of sleepers and rails in main tracks. Typically replacement investments extended over two or more track units. In those cases, replacement investment and minor repairs costs are divided into track units according to track unit's main track length. Track unit separation causes artificial cost dependence on track length, because track length is nearly same variable as main track length (the correlation between these two length variables is 0.97). Therefore, one additional guideline in track unit separation has been to have as few track units as possible, because then there is less need for artificial costs separation.

The data of independent variables gross tons, track length; number of switches and maintenance level is also based on the statistics of VR. The only exception of independent variables is speed, which is calculated from timetables. In every track unit, there were considerable speed differences between different type of train traffic. The variable used in this study is an average speed of different traffic categories.

Table 3 contains also an indicator variable K. There is not replacement investment about in 1/3 of track units. In addition, the average replacement investment cost is about 4-6 times higher than average maintenance cost, therefore there is clear difference in track unit's cost depending whether it had replacement investment or not. The indicator variable K models this difference (K_{it}= 0 or K_{it}=1 if track unit i at time t did not had or had over 0.1 million FIM replacement investment or minor repairs costs, respectively). Of course, indicator variable K is only used, when the independent variable is HLG-costs. In the case of regressing maintenance cost there is no need for variable K.

Table 3. Data for 1997, 1998 and 1999 (n=93)

Variable		Mean 1997	Mean 1998	Mean 1999	St.dev. 1997	St.dev. 1998	St.dev. 1999
Maintenance costs, million FIM	C_1	0.30	0.27	0.27	0.19	0.18	0.17
HLG- costs, million FIM	C_2	1.42	1.49	1.29	2.25	2.55	1.99
Gross tons (the natural logarithm)	Ln(U)	14.9	15.0	15.1	1.93	1.90	1.44
Tack length (km)	Y	80.6	80.6	80.6	45.1	45.1	45.1
No. of switches	z_1	45.0	45.0	45.0	31.9	31.9	31.9
Speed (km/h)	z_2	62.5	62.5	62.5	18.0	18.0	18.0
Maintenance level (1,…,6)	z_3	2.91	2.90	2.90	1.65	1.64	1.64
Renewal	K	0.81	0.86	0.86	0.39	0.35	0.35

3 The model

By specifying a cost function it is possible to estimate, how costs varies with explanatory variables (utilisation level, track length etc.). In this study, we have used similar expression like in P.Johansson and J.Nilsson's study. Thus cost function for track unit i at time t is:

$$C_{it} = g(Y_{it}, U_{it}, z_{it}, \varepsilon_{it}) = g(x_{it}, \varepsilon_{it}) \qquad (1)$$

C_{it} is independent variable, U_{it} is utilisation level (measured in gross tons), Y_{it} is the track length, z_{it} is a quality variable indicating for example number of switches or maintenance level and g is unknown functional form. In the study of Johansson and Nilsson, there was test for different kind of functional forms. Based on tests a Cobb-Douglas specification was chosen to be the most relevant functional form for the further analyses. A part from some modifications, we have used the same specification:

$$\ln C_{it} = \alpha_0 + \beta^y y_{it} + \beta^u u_{it} + \beta_z z_{it} + \beta_K K_{it} + \beta_K^y K_{it} y_{it} + \beta_K^u K_{it} u_{it} + \varepsilon_{it} \qquad (2)$$

Small letters indicate logarithmic form, here $y_{it}= \ln Y_{it}$ and $u_{it}= \ln U_{it}$. In the Cobb-Douglas specification β-coefficient are return to scale parameters. If the coefficient is under 1, it means that 1% increase in coefficient (gross ton) generates less than 1% increase in costs.

4 Estimation tests

The standard assumption of ordinary least squares (OLS) method is that variance of residuals are constant. However, the graphical examination with maintenance costs as well as HLG-costs shows that when plotting fitted values versus independent variables, the variance of residuals increases together with the increase in independent variables.

In other words, there are signs of heteroskedastisity. The calculated Cook-Weisberg test also indicated heteroskedastisity. To account for heteroskedastisity, estimation results are reported with the robust White's covariance matrix.

5 Estimation results

The results of the coefficient estimations are presented in subsection 5.1. In turn the subsection 5.2 reports the actual marginal cost of traffic volumes derived from the coefficient estimates.

5.1 The Cost structure

The results from the estimations are presented in table 4 and 5. In table 4 the independent variable is maintenance cost and in the table 5 HLG-costs. The obtained R^2 values indicate a considerably good fit for the models used. It is also notable that maintenance cost estimations gives higher values for R^2 than regressions with the HLG-costs. In other words the used model specification can predict better maintenance costs than HLG-costs.

The presented result tables 4 and 5 contain coefficient values only for statistically significant variables. In the case of maintenance costs this means that maintenance level and speed variables are not showed in the table 4. In addition, interaction terms K*u and K*y are excluded for the same statistical insignificance reason. There were also problems with multicollinearity. In turn the number of switches variable shown to be statistically insignificant when estimating the HLG-cost and therefore the number of switches is left out from table 5. The number of observation is also lower in estimation results tables than in table 3, which described the original data. This is due to the omission of the outliners.

By the estimations with the both cost category, maintenance and HLG-cost, the track length and the gross tons proved to be statistically significant. This also applies the indicator variable K, when estimating HLG-costs. The obtained coefficient values for gross tons are well below one, meaning the railway infrastructure maintenance and reinvesting to be a decreasing cost activity. In addition the comparison of tables 4 and 5 shows the gross tons coefficient values for maintenance cost to be lower than the HLG-cost ones.

The results from Finnish maintenance cost data are quite same as the results in Swedish reference study. The Swedish gross tons coefficient values for gross tons were also well below one, the values on main lines were between 0.13 and 0.28, and between 0.23 and 0.34 on secondary lines.

Table 4. Estimation results: maintenance costs

		1997 Coeff.	(t)	1998 Coeff.	(t)	1999 Coeff.	(t)
Gross tons	U	0.08	(2.90)	0.13	(6.37)	0.13	(5.36)
Tack length (km)	Y	0.60	(7.35)	0.61	(9.83)	0.65	(10.49)
No. of switches	z_1	0.39	(5.41)	0.42	(7.60)	0.34	(6.88)
Constant		9.54	(20.07)	8.61	(25.54)	8.67	(21.96)
Number of obs.		90		90		90	
R-squared %		80.6		86.0		86.2	

Table 5. Estimation results: HLG-costs

		1997 Coeff.	(t)	1998 Coeff.	(t)	1999 Coeff.	(t)
Gross tons	U	0.29	(4.86)	0.32	(5.41)	0.27	(4.43)
Tack length (km)	Y	0.95	(7.13)	0.77	(5.72)	0.90	(7.43)
Renewal	K	1.19	(7.15)	1.12	(5.61)	1.16	(7.25)
Constant		6.32	(5.68)	6.59	(6.00)	6.71	(6.86)
Number of obs.		91		91		91	
R-squared %		55.7		47.3		49.3	

5.2 Marginal costs

In the definition of marginal the guidelines of P.Johansson and J.Nillsson’s study is followed. The marginal cost (MC_{it}) per $Gtkm_{it}$ (gross ton kilometre) is calculated using formula:

$$MC_{it} = \hat{b}^{u} \frac{\hat{C}_{it}}{Gtkm_{it}} \qquad (3)$$

The formula gives a separate coefficient for all track units. Also in this study these coefficients showed scale economics i.e. the higher traffic volumes the lower the maintenance cost or the HGL-costs.

To get a single "average marginal cost "measure for the whole rail network the track activity weighted formula is used:

$$\overline{MC}_{it} = \hat{b}^{u} \frac{\sum_i \hat{C}_{it}}{\sum_i Gtkm_{it}} \tag{4}$$

Here $weight_{it} = Gtkm_{it} / \sum_i Gtkm_{it}$. This procedure generates the same level of revenues as if a separate charge is levied for each track unit.

According to Finnish Rail Administration statistics there were 32,89 (1997) 32,96 (1998) and 32,72 (1999) billion gross ton kilometres in the Finnish rail network. Using this information and "average marginal cost" coefficients calculated with formula (4) it is possible to calculate the revenues, that marginal cost pricing would give. The table 6 presents coefficients and respective revenues.

Table 6. Average MC-coefficients and respective revenues

Year	Coeff.	MC-Coeff. (0.01 FIM/Gtkm)	FIM (million)
Maintenance 1997	0.08	0.00064	21.0
Maintenance 1998	0.13	0.00094	30.9
Maintenance 1999	0.13	0.00096	31.3
HGL- costs 1997	0.29	0.00731	240.5
HGL- costs 1998	0.32	0.00747	246.3
HGL- costs 1999	0.27	0.00643	210.2

From the table above it can be seen that revenues gathered with marginal cost pricing based on maintenance cost are approximately ten times lower compared to revenues for the HLG-costs. But if the gathered revenue level is compared to the cost level used in the analyses, the difference between cost categories is smaller, namely maintenance cost based charging would cover 7-12 % and HLG-cost would cover about 17 % of the analysed cost, which are presented in table 2. In the Swedish reference study corresponding cost recovery values for maintenance cost were 10-12 %. The actual

From the table 6 is also seen that a minor change in average marginal cost coefficient causes a considerable alternation in gathered revenue i.e. every 0.0001 chance in coefficient value means over 3.2 millions FIM change in the gathered revenue. This means that a slight misspecification in data can have a great influence on the level of the charges.

6 Concluding comments

The analyses of Finnish railway maintenance and reinvestment costs shown that the main explanatory variables for these costs are track length and traffic volume (measured in gross tons). As expected, the obtained elasticity for marginal variations in traffic levels were well below one. This indicates that, if the railway infrastructure charges were based on the marginal cost estimations, the gathered revenues would recover only a minor part of the total cost. This also applied, when the renewal costs were taken account in the analyses.

The another conclusion was that the used Cobb-Douglas model specification could explain better the variations of the maintenance costs than the HLG-cost, which also include reinvestment cost. This is most likely due to the more stochastic appearance of the major replacement investments. In other words, the major replacement investment is made for a certain track section approximately once in every 30 year. For this study there was data available only for the three years. It is obvious that the ideal settings of analyses would have a longer time interval than three years.

References

Cook R. D. and Weisberg S. (1983): Diagnostics for Heteroskedastisity in regression, Biometrika 70: 1-10

European Commission (1988): Fair Payment for Infrastructure Use: A Phased Approach to a Common transport Infrastructure Charging Framework in the EU, White Paper, Directorate- General VII Transport, Directorate B- Inland transport

European Commission (1999): Calculating Transport Infrastructure Costs, Final Report of the Expert Advisors to the High Level Group on Infrastructure Charging (Working group 1)

Greene W.H. (1997): Econometric Analysis, 3rd Edition, Prentice-Hall, New Jersey

Johansson P. ja Nilsson J-E.(1998): An Economic Analysis of Track Maintenance Costs, Economics Department, School of Transport and Society, Dalarna University, Paper no 881 for the 8th WCTR

Workshop on Infrastructure Charging on Railways
Helsinki, 31 July – 1 August 2000

Swedish railway infrastructure charging in principle and practice

Per-Ove Hesselborn, Swedish Institute for Transport and Communications Analysis (SIKA)

1. The efficiency-oriented reform of 1988

As a result of the Swedish transport policy reform in 1988, rail infrastructure was separated from operations. *Banverket*, the government unit responsible for infrastructure, was then created from the integrated national rail company, SJ. Both the infrastructure unit (Banverket) and the operator unit (SJ) were, and still are, kept in public hands.

The main purpose of separating infrastructure from operations was to enable the application of socio-economic criteria for rail infrastructure investments and pricing. Thus, according to the new policy, investments in rail infrastructure should be based on CBA and rail charges should be set – by Banverket – so as to stimulate the efficient utilisation of the rail network. Efficiency may of course be further enhanced by allowing competition between operators. This possibility, partially utilised in actual policy, does not seem to have been emphasised by policy makers at the time of the reform however.

The purpose of this paper is to describe shortly current Swedish railway infrastructure charging policy and to comment on the practical difficulties involved with deriving and introducing efficient charges – that is, charges that may stimulate an efficient utilisation of the rail network.

2. The 1998 reform takes a further step towards efficiency – in principle

Achieving an efficient utilisation of the rail network requires, according to established theory, that charges are based on short run marginal costs (SRMC). The Swedish reform of 1988 did not, however, eliminate the prevailing cost recovery restriction, implying that operators could not be charged according to only SRMC. Therefore a two-part tariff system was introduced, with a variable charge equal to SRMC, and a fixed charge – an access charge – with the purpose of raising the extra revenue necessary for cost recovery. In reality, the fixed part of the tariff was set to match the annual tax paid by competing road vehicles (HGV:s) because, at the time of the reform, it was considered particularly important to harmonise competition between road and rail.

The cost recovery requirement has now, after the policy reform in 1998, been abolished. As a consequence, the fixed part of the rail charge has also been abolished. Thus, there is now a variable charge only. This charge should, in principle, be set equal to SRMC. Fixed costs should now be financed in the least distorting way possible.

The revised charging system introduced in 1999, characterised by variable charge components only, is supposed to reflect the short run marginal costs of train movements. However, after 1 July 2000, two new variable charge components with a financial purpose have been added – a “traffic information charge” and a charge to cover the fixed costs of the Öresund bridge. These are both expressed in terms of SEK per gross ton kilometre, thus presumably causing distortionary effects that are unnecessarily large.

3. The current variable charge and its relation to marginal cost

The variable charge consists of four components:

- A track charge,
- A marshalling charge,
- An emission charge and
- An accident charge

The role of the *track charge* is to charge for the marginal wear and tear cost caused by an additional train movement. This cost varies with factors characterising the infrastructure as well as the vehicles. Originally (1988) the track charge was differentiated with respect to both these dimensions.

To reflect the fact that a certain vehicle inflicts larger costs the worse the standard of the track is, a differentiation into two track categories was introduced. But a majority of the political voices soon found the differentiation objectionable on regional policy grounds and the differentiation was abolished after half a year.

A differentiation with respect to vehicle characteristics – weight, speed etc. – was introduced in 1988. This differentiation was abolished in 1999 since new studies carried out in 1997 and 1998 had shown that track damage could be satisfactory explained by the weight factor. As a consequence all vehicles now pay the same charge: SEK 0,0028 per gross ton and kilometre.

According to Banverket, an important reason for the limited variation in track wear and tear costs is that such variation is held back by differentiated speed regulation. Generally speaking, vehicles that inflict larger costs are given a lower permitted speed level.

Banverket has recently initiated studies to further investigate how track damage is related to various vehicle characteristics. The results of these studies may give support to a charge differentiation similar to the one that was abolished in 1998. According to Banverket, a differentiation in terms of track categories may be motivated on the same grounds that were referred to in 1988.

The *marshalling charge* was introduced to account for wear and tear occurring at marshalling yards. Today this charge is supposed to match the marginal costs typical for yards with certain automatic marshalling equipment. According to Banverket, the present charge – SEK 4 per vehicle – is probably too low for the relevant types of marshalling yards.

The present *emission charge* – SEK 0,31 per litre fuel consumed by diesel driven trains – reflects monetary values of NOx-emissions that are now obsolete. To reflect present valuations (used in CB-calculations within the rail and road sectors), a drastic raise of this charge is motivated.

Today's *accident charges* reflect calculated average accident costs for freight and passenger trains, respectively. For freight trains the charge is SEK 0,55 per kilometre and for passenger trains SEK 1,10 per kilometre.

4. Further attempts to estimate marginal costs

During the autumn of 2000 further attempts have been made to develop accurate marginal cost estimates. For example, Banverket and SIKA have started studies with the purpose of deriving more relevant estimates of accident costs as a basis for proposing a revision of the present accident charges. The focus is on accidents in level crossings. The purpose is to derive marginal risk changes and marginal accident costs caused by additional rail traffic for various types of crossings. Banverket and SIKA have also started work to investigate scarcity problems on the Swedish rail network in order to derive relevant scarcity fees. Finally SIKA has suggested a study to estimate marginal railway noise costs. This study is planned to start in 2001.

Railway Restructuring and the Preparation for the Introduction of Infrastructure Charging in Hungary

Dr. Katalin TÁNCZOS - Gyula FARKAS

Budapest University of Technology and Economics

Abstract

The top management of the company has worked out the reform program of the Hungarian Railway (MAV) with the contribution of independent experts and with the involvement of the Ministry for Transport and Water Management and the Ministry for Finance. The reform program has been submitted to and discussed by the Government. The decision about the railway restructuring was made at the end of the last year. The paper gives an overview of the reform program with special focus on the planned infrastructure charging system.

Introduction

The political and economic changes have had dramatic effects on transport market in the entire Central and Eastern European countries, among them in Hungary. The earlier very high freight transport intensity, expressed in tkm/GDP, decreased significantly because of the rationalisation of production and distribution. Instead of the high volume, heavy goods exchanged within the COMECON, products with high value and sensitive for time have been shipped. The new requirements for better quality and higher frequency and reliability of passenger transport services affected the railway sector particularly.

The road transport services were more flexible to respond to these new challenges. The new type of transport demand could be performed by the quickly privatised and liberalised road sector more effectively. The strong competition resulted in a significant decrease of the railway's market share in Hungary.

The situation of the Hungarian State Railways in the first half of the decade

The 130 million tons freight transport performance of MAV in the base year of 1989 decreased to 45 million tons for the middle of the decade and the stagnation became the main feature in the second part of the nineties. The changes in the passenger transport demand caused similar but not so dramatic decline for the railway in the country. The decrease of the performances in the same period was near to 30% but there was no increase in this number in the last 4 years. The competition from the individual cars is very aggressive; the road and the air sector have carried the growing demand of the passenger traffic.

The lack of domestic sources restricted the possibilities for the improvement of the railway services. Only the Hegyeshalom-Budapest railway line could get a quick reconstruction and this section had been developed for the 160-km/h speed. The other parts of the infrastructure and the rolling stock remained very deteriorated and costly maintained and operated.

In the early 1990s, the dramatic decline of the Hungarian State Railways, (MAV)'s freight performance led to a situation when its freight profit was insufficient to cover the gap of passenger financing (in 1994 and 1996, MAV had no freight profit at all to use for cross-financing), so the railway company was forced to get short-term loans guaranteed by the government to finance operation from Hungarian commercial banks. In 1994 and 1995, the government had to exempt MAV from the repayment obligations relating to loans for operation. Financing a part of passenger costs from guaranteed short-term loans proved to be the most expensive way of operation of railway passenger services.

Between 1989 and 1996, fares in real terms have increased by about 5%. However, the increase occurred before 1993 and, since then, fares have declined in real terms. Over the same period of time, MAV's passenger cost recovery ratio declined from 32% to 22%.

Steps toward the restructuring

The top management of the Hungarian Railways soon recognised the need to increase the knowledge of the market. They have made greater use of market research techniques and marketing methods and special marketing divisions have been created on the most important nodes of the railway network.

Parallel to the staff reduction the management started the rationalisation of assets (such as marshalling yards and rolling stock) and, since 1992, spinning off non-rail activities like forwarding, repairing railway vehicles and equipment, building and track construction, printing, design and gardening. As a result of this process, on 31 December 1996 MAV owned 100% of shares of 55 companies, more than 50% of further 8 companies and it was a minority owner of 41 companies.

The major steps of restructuring the relationship between the government and MAV were:

- The approval of the new Railway Act in 1993 and the establishment of Hungarian State Railways PLC on 1 July 1993.
- The Agreement between the government and MAV concluded on l February 1995 and renewed for 3 years on 20 January 1996 defining the partners' obligations towards each other, with particular regard to the financing of public service obligations and the separation of organisation and accounting of infrastructure from the commercial railway.
- The Government Decrees of 1994 and 1995 concerning the financial stabilisation and the rationalisation of activities of MAV resulting in the exemption of the railway company from the obligations relating to loans for operation.

Evaluation of the slow progress in reforming the rail structure

The content of the first phase of restructuring has been done until now has not fully cover the intentions of the Directive 91/440. The autonomous management of railway enterprises, separation of infrastructure from operation at least in the accounting system has been almost realised but old debts have been only partially cancelled. The open access to infrastructure has not yet been implemented.

The autonomy of the management is partially limited by the fact that the government being responsible for the improvement of railway infrastructure (and guarantor for large investment loans of the commercial railway) can decide on investment priorities based on political considerations rather than economic or financial rates of return.

The economic forecasts based on the TINA corridors show the future possibilities for the rail sector. The favourable geographical position of the country makes it possible to identify four transport corridors of the Trans European Network, which cross the area of the country. These are the next ones: the 4th, the 5^{th,} the 7th and 10th corridors. The long distance freight transport, the combined transport, the intercity services and suburban passenger lines with high traffic density will create an efficient market share for the rail if the quality of the services will be improved

Further steps of the railway reform to be done in the near future

The reform program has been submitted to and discussed by the Government. The decision about the railway restructuring was made at the end of the last year.

All the participants involved in the reform process recognised the necessity of a transparent system based on controlling. The planned holding system is the first step for the later disintegration of the supply side. It is already decided to separate the state owned infrastructure company which will be responsible for the development, maintenance and operation of the track and belongings from the commercial railway company, which will be responsible for the passenger and freight services, and will be the owner of the rolling stock. The later company will pay the track user fee, which will cover the cost of maintenance and operation, but the development sources of the infrastructure will be provided by the state.

Hungarian methodological issues of user charges for railway infrastructure

A key element of the restructuring process will be the defining and introducing of user charges for the railway infrastructure. Following the instructions of the 91/440 EU directive in some countries the effective taking apart has been already realised. Publications reported different results according to the different strategies applied in the different countries. Therefore all the published international methodological issues and available practical experiences have been taken into consideration at the preparation of working out of the Hungarian solution.

Analysing general the practice in Europe it can be proved that infrastructure managers take the real infrastructure expenses as a basis and calculate base charge on this basis – in compliance with the used service standard. The base charge is corrected with additional services, surcharges and they refund the innocent operators in the case of decrease of standard.

The Hungarian methodology uses the following rudiments:

- Railway track expenditure needed theoretically: technical cost (expenditure) standard (outlay requirements) of the railway track according to EU Directives and arrangements;
- Real technical standard of the track: comparing the ratio of the real level to former standards (norms);
- Base charge for railway infrastructure: paid by operators, corrected by different surcharges (region-developmental, business-political, furthering modern environment-friendly transport modes, etc.).

A charge-system must be worked out which is suitable for

- train, category of train;
- route;
- statistical section;
- line, part of network;
- category of line;
- total network

fee determination separately for passenger and freight transport.

On account of the „Rules to calculate costs of production in railway business" the costs of passenger and freight trains must be separately determined in order to avoid cross financing between the mentioned two main activities. The controlling approach is a very important factor, too. [8]

Fundamentally, the user charges for railway infrastructure can be determined by statistical sections and fees should be summarised at any level, route – separately for passenger and freight transport.

The fee is made up of a fixed and a variable price component, it is a two-tier system. The fixed component must be paid for the access to the infrastructure (access charge), and the variable component for the actual use made. The charges that must be paid by the users of railway infrastructure are composed of the following elements [9]:

1. Access charge - fixed
2. Base charge - variable
3. Additional charge - variable
4. Surcharge - variable
5. Reductions - variable
6. Repayments - variable.

Fee = Access charge+Base charge+Additional charge+Surcharge-Reductions-Repayments

The other items will be determined below.

Ad 1. Access charge

The access charge can be connected with a current costs, which are not depend on the actual use made and train-performance. This charge is composed of the following elements: administration costs of the IM, social and welfare costs, financial expenditures of the IM. Distribution of the current costs of the IM among railway operators are occurred on the basis of the number of train on the section (line, part of net or whole network) by the certain railway operator.

Ad 2. Base charge

The base charge can be calculated in knowledge of the technical standard of the lines and the corrected expenditure for the infrastructure. The corrected infrastructure costs include the real costs of the infrastructure (maintenance, depreciation and operational costs), the costs of traffic control, the technologically justified additional maintenance and depreciation, the costs of backward improvement and the quantifiable externalities.

The mentioned cost elements constitute the corrected expenditure normative demand of the infrastructure manager. If this claim is satisfied, then the infrastructure manager will provide the gradually improving infrastructure, even if the technical standard of lines is behind the EU-standards.

Table 1 shows the technical parameters and the factors which determine the quality of the railway infrastructure.

Table 1. Technical parameters and service standard determining factors

Technical parameters of the lines (Tp)	Factors which determine the service standard
Speed of line (km/h)	Linear scale
Axle load (ton)	Linear scale
Electrified line	Yes/No
Number of track	Single track; Double-track
Safety of line (Signalling)	Key interlocking
	Key wedging device
	Mechanical
	Mechanical with light signal
	All-relay interlocking (D55, D70)
	Electronic
Train stopping control	Mechanical with light signal
	All-relay interlocking (D55, D70)
	Electronic
Traffic control	Between stations
	Mechanical block
	Automatic block
Socio-economic value of line	Volume of passenger/freight transport

The AGC-agreement is fundamental for the improvement of the “A-category” national lines. The AGC-agreement contains the characteristics of the average European railway infrastructure. *Table 2* demonstrates the AGC-standards.

Table 2. AGC-standards for the European railways

	Speed of line	Axle load	Electrified line	Number of track	Safety of line	Train stopping control	Traffic control
Stat. Section	160 km/h	22.5 tonne	Yes	Double-track	All-relay interlocking	All-relay inter-locking	Automatic block

After comparing the Hungarian railway infrastructure with the AGC-standards the backwardness of the technical standard of the Hungarian Railtrack can be determined.

Before that comparison the service-package – which is offered by the rail track – should be analysed and to settle the factors of the service standard should be defined.

The passenger transport and the freight transport have different requirements towards the infrastructure manager, therefore certain factors of the service standard have to be separately determined. *Table 3* lists the factors of parameters, which specify the service standard of the passenger and the freight transport.

Table 3. Factors of parameters, which specify the service standard of the passenger and the freight transport

Factors which determine the service standard	Factors of certain parameters (%) (α_j)	
	Passenger transport	Freight transport
Speed of line	α_{1p}	α_{1f}
Axle load	α_{2p}	α_{2f}
Electrified line	α_{3p}	α_{3f}
Number of track	α_{4p}	α_{4f}
Safety of line	α_{5p}	α_{5f}
Train stopping control	α_{6p}	α_{6f}
Traffic control	α_{7p}	α_{7f}
Socio-economic value of the line	α_{8p}	α_{8f}
Total	100	100

Comparison of the national railway infrastructure (technical standard) with the European network is important in several respects:

1. As-is analyses (region and size of the backwardness);
2. Increasing of competitiveness of the Hungarian railway infrastructure (neighbouring countries, improvement of transit routes);
3. Drafting of the strategic goals (EU-corridors, increasing of the transport capacity and capacity utilization);

4. Access to the railway infrastructure and determination of the user charges for infrastructure;
5. How does service standard influence the change (increase/decrease) of the expenditures for infrastructure?

The AGC-agreement parameters can be considered as a basis of comparison base with the standard of railway infrastructure (common service standard). Condition of the national railway lines has to be assessed by considering the AGC-parameters, and after that the service standard of the statistical sections should be determined.

Table 4 contains "servicing factors (β_j)" (which are characteristic of the sections) and the scales (which belong to certain parameters)

Table 4. Servicing factors and scales

Speed of line (β_s)

β_s= *Section max speed/160 *100%*

Axle load (β_a)

β_a= *Section max axleload/22.5 *100%*

Electrified line (β_e)

β_e= *100%, if the line is electrified, otherwise* β_e= *0%*

Number of track (β_t)

β_t= *100%, if the line is double-track, otherwise* β_e= *0%*

Safety of line (β_{sa})

Signalling	Key interlocking	Key wedging device	Mechanical	Mechanical with light signal	All-relay interlocking	Electric
Servicing factors	β_{sa1}	β_{sa2}	β_{sa3}	β_{sa4}	β_{sa5}	β_{sa6}

Train stopping control (β_{tsc})

Signalling	Mechanical	Mechanical with light signal	All-relay interlocking	Electric
Servicing factors	β_{tsc1}	β_{tsc2}	β_{tsc3}	β_{tsc4}

<u>Traffic control</u> (β_{ts})

Traffic system	Between stations	Mechanical block	Automatic block
Servicing factors	β_{tsc1}	β_{tsc2}	β_{tsc3}

<u>Socio-economic value of the line</u> (β_{sev})

Socio-economic value of the line	Passenger transport (seat-km)	Freight transport (gross tonne weight)
Servicing factors	β_{sev1}	β_{sev2}

Total of the service standard of all statistic sections can be calculated (for the passenger and freight transport) with the following formula:

Total service standard $(Tss)=\Sigma_i (l_i * \Sigma_j Tp_{ij} * \alpha_{ij} * \beta_{ij})$, where
l_i – length of line i,
Tp_{ij} – technical parameter of the line,
α_{ij} – factor of standard (passenger and freight transport separately),
β_{ij} – factor of the parameter (passenger and freight transport separately).

The average service standard of the railway network can be defined with the next formula:

Average service standard $(Ass)=Tss/\Sigma_i l_i$, where
Tss – total service standard,
l_i – length of line i,

The theoretical base charge can be calculated in knowledge of the average service standard and the corrected expenditure for the infrastructure:

Theoretical base charge = Corrected expenditure for the infrastructure/Ass

After definition of the planned performance ratio (on the basis of train km) between passenger and freight transport the specific factors must be formed for the determination of the real base charge (separate own performance of the passenger and goods transport).

The specific base charge can be expressed by means of the theoretical base charge and the total service standard:

Specific base charge$_{passenger}$ = *Theoretical base charge*$_{pass.}$/$Tss_{pass.}$;
Specific base charge$_{freight}$ = *Theoretical base charge*$_{freight}$/$Tss_{freight}$

Their specific base charge must be re-multiplied by the real own performance of the passenger and goods transport and after that the base charge (passenger and freight separately) can be defined to the statistic sections, levels, routes and network.

$$\textit{Base charge}_{section\ p,f} = \textit{Specific base charge}_{p,f} * l_i * \Sigma_j\ Tp_{ij} * \alpha_{ij} * \beta_{ij}\ ;$$
$$\textit{Base charge}_{route\ p,f} = \Sigma_i\ \textit{Specific base charge}_{p,f} * l_i * \Sigma_j\ Tp_{ij} * \alpha_{ij} * \beta_{ij}$$

The base charge includes the mandatory services. It has overriding importance, because of the insurance of the assuring the operation of the railway network. The mandatory services comprise guarantee of traffic safety, giving assistance in the case of accident, dangerous and special goods transport.

On payment of the base charge the infrastructure manager provides the following services: access to the tracks, stations, public freight facilities, yards/depots, basic information to the public and assistance in case of accidents.

Ad 3. Charge of additional services

It includes services, which are needed for secure railway transport (traffic control, traffic security), the usage of the catenary, refuelling and the use of different services at the stations, yards, warehouses.

Ad 4. Surcharges

Extra charges is justified in the following cases (over the base charge):

1. Additional trains on the heavy traffic lines;
2. Use of Budapest's terminal for passenger trains;
3. Exceeding of the punctuality needed (set down in the services contracts);
4. Access to information and communication network;
5. Non-use of booked paths;
6. Depending on whether when the train is running (day, period of day).

Ad 5. Reductions

Infrastructure manager can give reductions in the following cases:

1. Taking advantage of the path for more years – on the basis of the contracts;
2. Use of the weakly frequented lines;
3. Running train at the same time (on all days of the week);
4. Shifting paths from saturated lines to alternative lines/services;
5. Shifting paths from peak to off-peak hours.

Ad 6. Repayments

If the infrastructure manager does not perform or incompletely complies with its obligation (that it was undertaken in the "services contract"; e.g. infrastructure failure, bad signalling or plan delays in all services) then railway operators can claim a penalty. If a railway operator's train breaks down and delays other services, then he pays penalty to the infrastructure manager and after that the railtrack pays penalty to other (innocent) operators. If the services run better than benchmark due to good network performance then the operators pay bonus to the infrastructure manager.

Summarising the elements of the user charge it can be mentioned that the fee is an index number, which expresses the quality of the railway infrastructure. The value of the charge shows the changes of the service standard. The user charge for the railway infrastructure involves a direct feedback and they show how the increase/decrease of the state subsidy influences standard level of the rail transport.

The user charges have to be determined so that they demonstrate the condition of the national railway infrastructure – in comparison with EU-railways – and inform the national/foreign railway companies should be informed about the volume of the fee.

Finalising the preparation phase of the railway reform, the followings have to be done in the near future:

- adjustment to EU Directives;
- the methodology should be implemented;
- opening of the Hungarian rail transport market.

Conclusions

An intensive work has been started to make the clear distinction between the loss-responsibility of the state and the MAV. After the careful separation of the loss-elements the burden of the central budget will decrease.

The reform does not mean the provision of more money for the MAV but controversially, its aim is to create competition and efficiency. The size and the network of the railway will be fit to the demand, the assets will be operated with higher efficiency, and the staff will be used with more responsibility and stronger motivation.

Among the future plans is the introduction of new competitive services, new type of co-operation with other modes in providing combined transport. Implementing new logistics centres the Railway Company wishes to further extend its links to the clients.

In the passenger transport the aim of the Hungarian Railway is to develop and offer long-distance services and will focus on suburban traffic. The company wishes to provide excellent conditions on the trunk network to help the joining process of the country to other counties of the European Union.

References

Rail Network Co-operation in the Age of Information Technology and High Speed. ECMT, Paris, 1989.

The Separation of Operations from Infrastructure in the Provision of Railway Services. ECMT, Paris, 1997.

Tanczos, K.: First signs of growth on Eastern Europe's railways. The World Railways.1996/97

Hager, W. – Rothengatter, W.: The prospects of the railways in the CEE countries, CEPS Working Party. September 1999

Salavecz J.: Information on development of user charges for Railway Infrastructure in Hungary. Seminar Bratislava, 7-9 July, 1999.

Access issues in Latin America railroads: what could Europe learn/avoid?*

Javier Campos**

University of Las Palmas

August, 2000

Abstract

This paper describes in detail the recent rail restructuring processes in Latin America (particularly, in Mexico and Brazil) focusing the discussion on infrastructure access and charges. It is shown that the spirit of all these changes has been decentralization and disintegration, together with increased private sector participation. Our approach follows that spirit, pushing the reforms one step further and drawing positive and negative lessons to be learned/avoided by European countries. On the positive side, we should include the relatively smooth way in which private participation has been introduced in a sector where many countries thought at the beginning that it was not possible. A second positive message is the way in which access issues and charging principles have been setup in concession contracts. The most relevant negative lesson, the one European countries should possibly avoid, is the need of shorten the transition period and strengthen the political compromise, thus limiting the short-term reversibility of the process.

* The views expressed in this paper do not necessarily reflect the opinion of the institutions the author is currently or has been affiliated with and, in particular, they do not necessarily reflect the opinions of The World Bank or the University of Las Palmas. The author gratefully acknowledges comments and useful discussions from participants at the Railroad Conference held in Paris (July 2000) and the Helsinki Workshop on Infrastructure Charging (August 2000). The paper has been partly financed by the Spanish Ministry of Education, CICYT Project SEC99-1236-C02-02.

** *Contact address:* Department of Applied Economics. University of Las Palmas. 4, Saulo Torón. 35017 Las Palmas. Spain. Phone: +34928451792. Fax: +34928458183. Email: jcampos@empresariales.ulpgc.es.

1. INTRODUCTION

The restructuring of the main industries that underpin every economy providing an environment in which economic growth can occur is at the heart of the market reforms that have swept the developing world during the 1990s. This is particularly the case of transport infrastructures and services, two sectors where a number of important changes have occurred – and are still happening – in Latin America, Eastern Europe and Asia.

During the last decade, there has been a dramatic increase in the liberalization of transport policies and a strengthening of the role of private operators and investors in transport infrastructure and services in these countries. This increased private sector participation has often reflected changing ideologies about the role of the state and dissatisfaction with publicly provided services. However, the main driving force behind it has generally been the pressure to look for private financing imposed on governments by lasting fiscal crises. This change in the financing of the sector is also providing an opportunity to restructure the transport industry in an attempt to improve its efficiency and sustain these improvements.

Focusing on Latin America, the purpose of this paper is to draw core lessons from the experience of two countries in the region, in order to provide new and old governments in the region and elsewhere with better information related to how they could structure a reform package in transport to make the best of the growth opportunities within their countries. This should help bring about the economic growth that is central in helping to alleviate poverty in developing countries and moving these economies out of the stagnation that they face. On the way to achieving this aim, some useful lessons could also be extracted for European countries.

Latin America is a good example to base this paper on because most countries in the region display many of the social and economic problems experienced throughout the developing world, such as significant migration from rural to urban areas and the consequent need for rapid expansion of service delivery combined with low levels of per capita income. In addition, within Latin America there are countries that have been at the forefront of reform. Chile was among the first in the world to undertake significant reform of the railroad industry and created some innovative methods for road concessioning; Argentina's overall transport sector reform is viewed as an example in several other places, and Bolivia's capitalization program is a model of whole-scale reform of infrastructures. Finally, while the reforms have, overall, been successful, there are also lessons of how well intentioned reforms can have a negative impact on growth.

This paper focuses on the rail sector, where some changes have been more radical than others, and where there is a generalized element of experimentation. Yet the spirit of all of them is decentralization and disintegration, with increased private sector participation. Our approach follows that spirit, pushing the reforms one step further and drawing positive and negative lessons to be learned/avoided by European countries. On the positive side, we will include the relatively smooth way in which private participation has been introduced in a sector (rail infrastructure and services) where many countries thought at the beginning that it was not possible. A second positive message will be the way in which access issues and charging principles have been setup in concession contracts, showing that – in principle – these complicated issues can be addressed through ex-ante negotiations. The most relevant negative lesson, the one European countries should possibly avoid, is the need of shorten the transition period, thus limiting the possibility of political interference over time.

To address these issues in order, after a brief introduction on the Latin America transport restructuring process (Section 2), we will use the examples provided by the cases of Brazil (Section 3) and Mexico (Section 4) to illustrate the main points. We finally conclude in Section 5 with a general discussion of the lessons learned from these countries and extrapolate its possible consequences for Europe.

2. TRANSPORT REFORM IN LATIN AMERICA

2.1. Some characteristics of the Latin America transport system

From a historical perspective, Latin America transport system can be viewed as a network of traffic corridors where – with the exception of the landlocked countries, Bolivia and Paraguay – the movement of freight and people always started and ended at the main ports in the Atlantic and the Pacific, and was only gradually penetrating the richer areas in the interior.[1] The centralized economy and society that the Spanish and Portuguese colonial authorities promoted during the centuries that followed the first European settlements favored the development of single corridors from production sites to export ports and paved the way the railroads would use in the 19th century.[2] It was only after the World War II when roads became the dominant transport mode, by connecting growing urban industrial sites with the coast and the interior.

As illustrated by Table 1, during the last forty years the kilometers of roads have been growing at a faster pace than railtrack has been declining. In 1960, the ratio of kms of paved road per km of rail line was 0.64, whereas in 1990 the same figure had risen to 3.35. In some countries, such as Brazil or Argentine, the reduction of the railtrack was above 20%. Simultaneously to this physical decay there was a substantial fall of the market share in both freight and passenger markets during the 1970s and 1980s, which apparently stabilized during the 1990s in the freight market. This decline was particularly relevant because it took place in a period when the total volume carried in both markets experienced a growth of about 50% in the region. Thus, the rail industry appears not to have been able to take advantage of the growing demand for transport in the last four decades.

---

[1] Of course, this 'colonization development model' was not particular to Latin America. However, the most notable difference with the United States or Canada was that the transport network was less dense and more focused on carrying goods to large distances, rather than people.

[2] In fact, the development of modern transport systems, such as the railroad, was faster in Latin America than in mainland Spain and Portugal, and even faster than in other European countries. However, these pioneering experiences were highly output-dependent routes, and when agricultural and mining crisis hit the region, many corridors were closed or abandoned.

Table 1. Roads and railways in Latin America (selected countries)

	Paved roads (kms)				Railtrack (kms.)			
	1960	**1970**	**1980**	**1990**	**1960**	**1970**	**1980**	**1990**
Argentina	22,712	33,375	52,194	57,280	43,905	39,905	34,077	35,754
Brazil	12,703	50,568	87,045	161,503	38,287	31,847	28,671	22,123
Chile	2,604	7,411	9,823	10,983	8,415	8,281	6,302	7,998
Colombia	2,998	5,980	11,980	10,329	3,161	3,436	3,403	3,239
Mexico	25,667	42,674	66,920	82,022	23,369	24,468	20,058	26,334
Total region	85,514	182,088	267,962	370,059	132,470	120,045	105,691	110,301

Source: The World Bank (1995). Total region also includes Bolivia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Venezuela and Uruguay.

This substantial reduction in market share, which is not only particular to Latin America countries but also a common trend around the world, can be attributed to both exogenous and endogenous causes. The first category includes the rapid development of alternative modes of transport, especially by road. In freight transport, an expanding and competitive trucking sector gained a growing percentage of road transport in many countries. The endogenous causes of the decline could be summarized in the inability of the sector to adapt itself to the changing conditions of its economic environment. Regulation remained obsolete and the rail industry was slow to react. The policies adopted during the 1970s and 1980s did not halt the steady loss of market share, the growing financial deficits, and in some countries, the impossibility of raising the low productivity indices of the industry. Thus, more radical restructuring processes and overall reforms were put into practice.

2.2. The need for institutional reforms in transport

From the Great Depression onwards, many Latin American countries had adopted state-led development models. Governments not only regulated most socio-economic activities but also became the largest producer, employer and consumer in many of these countries through the establishment of special agencies and state-owned corporations. Rail transport was not an exception and, at the beginning of the 1980s, most services and infrastructures were still controlled by the state, either at the regional or the central level. This situation was a direct result of the development models of the 1950s and the 1960s, and was partly based on the narrowness of the private sector initiative. In countries like Venezuela, Brazil or Mexico, public sector expansion programs were directly responsible for the high rates of growth of these decades.

Infrastructure privatization – and particularly in the case of the rail industry – was not likely to take root, succeed and achieve its public policy objectives under these circumstances unless private investors could be assured that their investments would generate an adequate economic return. In the rail industry, where deterioration under public hands was evident, urgent reforms were needed in order to encourage and sustain efficient private sector participation.

This change in the countries' economic framework was not easy and had to include several key aspects at least. The first one was to clearly define the scope and nature of private property rights in the provision of public services. Second, this framework should also encourage the efficient flow of private resources to infrastructure by removing obstacles to private provision of services. The provision of public sector services that were not privatized outright also required a clear environment for efficient contracting between the government and the private sector. Finally, the legal and regulatory framework associated to these changes should also provide an efficient mechanism for resolving disputes between the parties and assure the private parties' recourse to fair and speedy dispute resolution mechanisms.

The rail industry in countries such as Chile, Argentina, Bolivia, Peru, Mexico or Brazil embraced these reforms with more or less forced enthusiasm. In most cases, however, even the constitutions or the sectoral laws limited the participation of private sector, not only in the ownership but also in the operation of services, which were attributed to the exclusive responsibility of the existing public corporations.[3] Private participation was often subject to the will of the government authorities. This restriction stemmed from the fact that some countries considered "public services" to be the exclusive domain of the state. Under these circumstances, the alternative of outright privatization of services and infrastructures was always a difficult one. Important legal amendments and constitutional changes were needed to sell state-owned enterprises. Argentina and Chile had to change their entire legal systems, Brazil passed a new constitution in 1988; Mexicans had to reform theirs in 1995. Therefore, alternatives without full privatization – particularly those involving concession of rail services – were preferred. Chile and Argentina started this movement in the mid-1980s, but the processes where the reform was taken more extensively, and where access issues were more relevant, are the experiences of Brazil and Mexico.

3. THE REFORM OF THE BRAZILIAN RAILROADS

3.1. The Brazilian rail industry

The first rail line in Brazil was completed in 1854 by private foreign capital. During most of the following 100 years, private operators dominated the industry, but with an increasing participation of the public sector. In 1957, culminating nationalization policies of previous years, Federal Law 3115/1957 was enacted, incorporating under the jurisdiction of the Ministry of Transport the government-owned Federal Rail Network Corporation (RFFSA or *Rede Ferroviária Federal, Sociedade Anônima*). Twenty years later, a second operator in the form of a state-owned corporation, *Ferrovias Paulistas, Sociedade Anônima* (FEPASA), was created by State Law 10410/1974, which also established rules for the state of Sao Paulo financing of uneconomic rail services, absorbing the contributions to the workers pension fund, and other liabilities of the existing operators within that State.

[3] For example, *Ferrocarriles Argentinos* in Argentina, RFFSA in Brazil, or *Ferrocarriles Nacionales de México* had the exclusive right to provide rail services, according to their respective sectoral laws.

These two operators provided rail transport services to about 95% of the country's freight shippers, whereas the third important operator (and the largest in terms of output at the beginning of the 1990s) was the *Companhia Vale de Rio Doce* (CVRD), a huge government-owned industrial holding that exploited two specialized rail lines, EFVM (*Estrada de Ferro Vitória a Minas*) and EFC (*Estrada de Ferro Carajás*), from their mining sites to the ports in the north and center of the country. This company only served its own traffic, which mostly consisted of large volumes of iron ore for export.

By 1996, several restructuring procedures had been attempted to tackle the most urgent needs of the industry while maintaining it within the public sector. These policies, however, were not enough and the government started to look at the successful experiences of Argentina and Chile. Encouraged by these examples, Decree 473/1992 included RFFSA in the Brazilian National Privatization Program in a political movement that represented the first major privatization of public infrastructure services in Brazil. At this moment, in view of the geographic characteristics of the country, the size and state of conservation of the railway network, as well as the significant cross-regional differences in traffic, it was decided that the restructuring process could be more easily implemented if based upon RFFSA's existing regional structure.

Table 2. Economic characteristics of RFFSA concessions[4]

	Oeste	Centro-Leste	Sudeste	Tereza Cristina	Sul	Nordeste	Paulista
Concessionaire	*FNV*	*FCA*	*MRS*	*FTC*	*FSA*	*CFN*	*FEPASA*
Track Length (km)	1,621	7,080	1,674	164	6,586	4,534	4,236
Track Gauge (m)	1	1	1.6	1	1	1	1.6 & 1
Locomotives	88	397	406	10	395	112	408
Wagons	2,777	9,233	11,406	563	10,626	1,919	11,855
Output							
In 1995	1.6	6.26	20	0.10	7.5	0.7	6
2002 (est.)	5	26	37	0.16	24	4.4	17.2
Operat. revenue							
In 1995	37	175	321	8	187	26	187
2002 (est.)	86.0	350.0	490.0	9.7	327.0	70.3	243.1
Employees	2,423	10,982	9,397	343	9,604	3,707	13,432
(Transferred)	(1,800)	(7,900)	(6,600)	(250)	(6,900)	(1,600)	(6,380)
Main Cargoes	Petroleum Soybeans Steel Minerals	Petroleum Cement Steel Soybeans Grains	Iron ore Cement Steel Limestone	Coal and by-products	Soybeans Petroleum Rice Alcohol	Iron ore Petroleum Oil Cement	Petroleum Oil Minerals Grains Pellets

Source: Campos and Alexander (1999).

[4] Figures for track and rolling stock and employees correspond to 1998 actual values. Figures for output (in TKU billion), operating revenues (in US$ million) correspond to 1995 (before the concession) and to the estimates for 2002, the sixth concession year.

RFFSA's network was separated into six vertically integrated monopolies (called *malhas*) whose rail services would be concessioned out by the Ministry of Transport, and whose rolling stocks and existing infrastructures would be simultaneously leased by RFFSA to the private operator. The reason for this double concession-leasing method was that, according to the 1988 Constitution, the federal government had to remain the titular to the right of providing rail transport services in the country and, in addition, retain under its ownership the assets involved in those services.

As shown by Table 2, six concessions – *Nordeste*, *Centro-Leste*, *Sudeste*, *Sul*, *Teresa Cristina* and *Oeste* – were awarded between 1996 and 1997. Four of these railroads connected ports along the coast with their respective hinterlands, approximately 400 kms inland. On December 23, 1997, FEPASA was transferred to the federal government and in May 1998 the *Malha Paulista*, as it was also known, was immediately included in the privatization program. Its sale took place in November 1998 and concluded the privatization process of former government-owned rail operators.

Finally, when CVRD was privatized in June 1997 its two railroads (EFVM and EFC) were sold with it as part of the industrial holding; they were not concessioned in the same way as the RFFSA network. Since they had been originally designed to connect the company's mines and mills with one another and with the exporting ports of Vitória, Tubarao and Sao Luis, the railroads were kept with the company under control of the new owners. The two railroads essentially now operate as internal departments of CVRD, specialized in iron ore traffics, although they are obligated to carry traffic for other shippers as well.

3.2. The concessioning process

Except in the case of CVRD, the concessioning was implemented through public competitive bidding for the operation and maintenance of each of the *malhas* for a period of 30 years (renewable for another 30 years at most) with the simultaneous leasing of operational assets by RFFSA and the sale of some small non-operational assets. There were no prequalification requirements for candidates and the only limit established to avoid excessive concentration of ownership was that the share of each economic group participating into a concession should be limited to a maximum of 20% of total stock. However, no restrictions were imposed for cross-participation in different concessions or about the participation of major rail users, clients or suppliers as shareholders in privately operated concessions.

Each auction was won by the highest bid consortium, whose bid had to be above a minimum stipulated by the government. The amounts paid by each concessionaire – a down payment of between 10-30% of the minimum price and quarterly installments for the rest – were shared by the Federal Treasury (5%, corresponding to the concession of rail services) and RFFSA (95%, corresponding to the lease of assets). Five of the seven RFFSA concessions sold for more than the minimum bids. This success was due in part to the fact that the government reduced the workforce by approximately half in advance of the concessioning (see Table 2), and also in part due to the relatively stable macroeconomic environment during these years. The government had to receive a total of about R$1,700 million (US$950 million) for the seven concessions, although only about R$400 million was paid in the first installments with the rest due (after a 1-3 year grace period, depending on the concession) in 108-112 quarterly payments over the remaining life of the concessions.

There were no specific investment obligations set in the contracts. They only spelt out two specific targets on output and safety, in terms of minimum net ton-kilometers carried each year and maximum number of accidents per train-kilometer during the first five years. These targets would be reviewed during the third concession year, establishing the new goals for the next five-year period. Although it was clearly indicated that the reviews would be based on substantiated studies of past performance, the process had not yet started in 1999. The implicit idea behind these targets was that in order to meet them the concessionaires would have to carry out investments and therefore they all were obliged to submit in advance a triennial investment plan to obtain clearance.

With regard to the relationship between the concessionaires with the final users, the maximum prices to be charged for transport services were also set in the contracts. Ceilings varied according to the length of the haul, type of product and the geographic region served. These prices were to be periodically revised to correct them according to inflation. There also existed a vague notion regarding the concessionaire’s obligation to maintain its financial and economic equilibrium: the concession contract determined that the tariffs should always be above the railroad long run variable costs, although no methodology was provided for the calculation of these costs.

3.3. Issues related to access among concessionaires

Since the payment to RFFSA for the leased assets (including tracks) was clearly specified in the contracts, the most important access issue that faced the Brazilian rail industry at the moment of the privatization was the development of cross-concession traffic. Traditional cargoes, such as ores, iron and steel, needed to travel from inland to the main cities and seaports, i.e. east to west. New products, final goods and half-elaborated commodities, however, were creating an increasing need for north-south traffic, particularly to access the cities of Sao Paulo and Rio, and their respective ports. Since the concessions were let on the basis of the old approach, north-south traffic needed to cross several concession areas.

Establishing a process to allow this cross-concession traffic was identified by all the players within the industry as a key issue to be addressed soon, so that the growth potential of these new markets could be unleashed. Thus, the Brazilian concession contracts included special provisions for captive shippers, joint traffic and access rules to other networks. In general, it was expected that the interested parties would reach an agreement on these issues. If not, the government, through the Ministry of Transport, had the power to review the problem and set rates for shippers that were captive to the railroad. Railroads were also obligated to carry joint traffic or, if they could not, to allow the connecting railroad access to its tracks so that it could complete the movement. The two railroads were to negotiate the tariffs for joint traffic, but again the government could step in to set the rates or order access if the negotiations failed.

Although the restructuring model chosen for Brazilian railroads intended to minimize access issues by reorganizing the industry into separate and relatively little interconnected networks, the horizontal separation model made it clear that in most cases a concessionaire would have to use its neighbors' tracks when carrying long-distance traffic. In principle, the government did not worry excessively about this issue during the privatization of RFFSA because the six RFSSA companies that were formed interchanged little traffic with one another. But they interchanged with the CVRD railroads and with FEPASA, and the privatization of FEPASA in 1998 (and particularly the access to the port of Santos, which included an internal rail network of about 200 kms) brought the issue of joint traffic to the forefront again.[5]

In fact, according to available figures, despite the historical lack of connection among several lines due to distance or different gauges, the interconnectivity issues were very important in 1995, particularly in central and southern Brazil. However, some railroads were much more dependent on joint traffic than others. For example, as shown by Table 3, of the 17 million tons originating on FEPASA, almost 2.5 million tons was transferred to MRS, mostly to be shipped out of the port of Santos. From MRS' perspective, however, this cargo accounted for less than 10% of MRS' total tonnage, and even less of its ton-kilometers because they traveled only the last 22 kilometers or so of their journey on the MRS system.

Table 3. Interchanges of traffic in 1995 ('000 of tons)

Destination / Origin	CFN	FCA	MRS	FSA	FNV	FTC	FEPASA	CVRD	Total	% ending on own line
CFN	1,862	68	-	-	-	-	-	-	1,930	96.5
FCA	14	11,558	2,573	-	-	-	794	2,589	17,528	65.6
MRS	-	77	43,850	-	-	-	2,928	399	47,254	92.8
FSA	-	-	-	14,460	-	-	791	-	15,251	94.8
FNV	-	-	-	-	1,736	-	727	-	2,463	70.5
FTC	-	-	-	-	-	1,336	-	-	1,336	100.0
FEPASA	-	914	3,869	1,597	928	-	11,814	-	19,112	61.8
CVRD	-	2281	600		-	-	-	84,913	87,794	96.7
Total	1,876	14,898	50,892	16,057	2,664	1,336	17,054	87,901	100	-
% originated on own line	99.3	77.6	86.2	90.1	65.2	100.0	69.3	96.6	-	-

Source: Campos and Alexander (1999).

[5] The port of Santos – the most important maritime port in Brazil – constitutes the hottest access issue at the moment. During 1998-99, there was an important conflict between rail operators and the port administration (*Codesp*), but since January 2000, an agreement for joint operation of the internal network has been reached. More recently, in July 2000, the port authority formally transferred the control of the tracks to the private firms, which now jointly operate the internal network. Again, access prices and rules are set by bilateral contracts, without government intervention.

In 1999, complaints about joint tariffs were common among the carriers. FEPASA and FSA were in a tough dispute over through rates, for example, and FEPASA regarded the rates that MRS charged for access to Santos as excessive. However, no complaints had been brought to the regulatory agency so far, which suggests that the railroads were still hopeful of negotiating reasonable solutions without appealing to external control mechanism.

As mentioned above the general policy set in the contracts on access rights, mutual traffic, multimodality, etc. favored bilateral, market-based solutions, giving again only the power of arbitrage to the Ministry of Transport. This implicitly reflected the idea that if regulatory authorities prevented the abuse on captive shippers then there was no need for them to also regulate the division of tariffs for joint traffic or to order one carrier to allow another access to its tracks. In theory, such matters could be handled by negotiations among the carriers, much as they would be in a normal competitive market. However, in the case of Brazil, this approach required two prequalification criteria hardly met: that the regulators were able to regulate tariffs for captive shippers and that the railroad management was experienced and sensible about negotiating joint tariffs. This final point required the management of the concessions to have a single objective, that of profits, which was not necessarily the case of Brazil.

4. THE REFORM OF THE MEXICAN RAILROADS

4.1. The Mexican rail industry

Railroads began operations in Mexico in the late 19th century, when several US companies used imported materials to build lines along the country's Pacific coast. The largest company, *Ferrocarriles de México* (FdM) became Mexican-owned in 1908 and was later nationalized in 1937. In the 1980s FdM and the remaining rail lines were incorporated into *Ferrocarriles Nacionales de México, SA de CV* (FNM), controlled by the Transport Ministry (*Secretaría de Comunicaciones y Transportes*, SCT), and in 1983 the Constitution was amended to formally require that the federal government owned and operated all main railway services in the country.

FNM was an integrated monopolistic railroad that provided freight services in both the national and international markets. It also provided some inter-city passenger services, but did not supply any commuter passenger services to Mexico City or any other major city. In 1996, the overall system was composed of 26,623 kms of track, of which 77% were primary lines divided into three main geographical divisions, Pacific-North, Northeast and Southeast. The remainder formed the short lines, the network that served the metropolitan area of Mexico City and some small private lines. Like many other state-owned rail companies, FNM had developed a production-oriented, rather than a commercial-oriented culture. Although some of its operating performance indicators were comparable to those of similar countries in the region (for example, average hauls of 2,830 tons, train-lengths of 41 cars, etc.), others clearly reflected several sources of inefficiency (for example, average train-speed was only 25 km/h and average daily distance traveled by locomotives was below 250 kms) related to the age and state of maintenance of track and rolling stock. Safety concerns related to the number of accidents, spoiled cargo and theft were also high.

4.2. The reform of Mexican railroads[6]

As early as 1980, the Mexican government was aware of the medium-term evolution of the rail sector and reckoned that its poor performance hindered the development of the country. In the period from 1982 to 1989, several institutional reforms within the existing system were attempted but they failed. President Salinas' administration (1989-1994) was marked by more significant improvements in performance but also by a calculated ambiguity about whether the railroad might eventually be privatized. In 1992, a new Director General for FNM was appointed and he announced a Program for Structural Change (*Programa de Cambio Estructural*, PCE) whose main goal was to establish a more commercially oriented railroad.

The plan was designed to enhance the company's efficiency and productivity by focusing on freight transportation as the core business and eliminating some unprofitable services. Arguably, the most important PCE reform involved labor: the workforce was reduced from approximately 80,000 to 50,000 employees, largely through a program of voluntary retirements. With the unions' cooperation, moreover, the book of work rules, which had been unchanged for many years, was simplified and modified to increase labor and locomotive productivity. Under the PCE the financial performance of the railroad also improved, but not enough to reverse the trend of previous years.

When President Zedillo took office in December 1994, the Finance Ministry was reportedly disappointed with the rate of improvement under the PCE. The pace of the restructuring process was accelerated and, in February 1995, the Mexican Congress approved a new amendment to article 28 of the Constitution, which reclassified railroads as a priority activity, thus opening opportunities for private sector investment within the railway system. In May 1995, the *Ley Reglamentaria de Servicios Ferroviarios* (LRSF), a new sectoral law regulating railway services, outlined the general procedures for these investments and defined the conditions under which private participation in railways was going to be allowed for the first time in forty years.

After discarding some alternative proposals, the scheme chosen for privatization involved the geographical separation of FNM's assets and operations to setup a number of route-based companies according to the pre-existing regional divisions. Each of these companies was awarded a 50-year *concession title* describing service conditions and overall relationship with the federal government and other private operators. The concessions could be extended for up to an additional 50-year term and, in general, they allowed to operate, exploit and, if required, build new lines with the goal of providing public railway transportation and ancillary services specified in their respective titles, real property, facilities and other equipment required for the operation of the company and certain liabilities. Under this format, vertical integration of the different functions or services in FNM was preserved, although functions could be unbundled whenever it was deemed necessary.

[6] This section draws on Gomez-Ibañez (1997).

After the horizontal breakup, the final stage of the privatization process was the sale of the shares owned by the government in the concessionaire companies through a public bidding process open to private investors. The government decided to auction first 80% of the shares of the capital stock of each of the companies and retain a 20% stake in each with certain limited rights. The government also obliged itself to sell the remaining 20% stake in each company through public offerings within five years of the disposition of the relevant 80% stake.

Table 4. Mexican rail concessions

	Pacific-North	North-East	South-East	Short-lines
Track (as a % of total)	30.3	19.3	10.7	38.7
Freight traffic (as a % of total)	46.2	37.6	8.6	7.8
Revenues (as a % of total)	44.7	37.1	9.8	8.4
Main cargoes	Iron, coal, oil	Corn, wheat, iron	Corn, wheat, oil	Vary across regions
Major industrial cities	Mexico City Monterrey Guadalajara	Mexico City Monterrey Guadalajara	Mexico City	Several
Major ports (*)	Tampico (G) Manzanillo (P)	Tampico (G) Veracruz (G) Laz. Cardenas (P)	Veracruz Coatzacoalcos Salina Cruz	None

Source: SCT (1996). *Notes:* (*) P= Pacific; G= Gulf.

The overall privatization scheme recognized that the main demand for rail services in Mexico came from freight carriers. With respect to passenger transport, apart from those lines already included in the concessions, several services would be privatized by assigning the concessions to companies bidding for the lowest subsidy. This process would only be applied to routes that lacked an alternative transportation mode. In other cases, passenger services would simply disappear since road transport was perceived as a generally adequate transport means for the country.

It is important to note that the 1995 railroad law (LRSF) kept the regulation of the privatized Mexican rail industry after the auctions within the SCT, particularly under the control of the *Dirección General de Tarifas* (DGT), a 250-staff regulatory body who was also in charge of tariffs (other than in railroads) and multimodal issues. However, as compared to the period when it also ruled over FNM, the regulatory functions of this body were now limited to supervise the activities of the concessions, devise the general policy for the industry and act as a arbiter in case of conflict among concessionaires.

According to the concession titles, the concessionaires were free to set their own tariffs in recognition of the extensive competition from trucks and the potential for competition among the concessions. Maximum prices were registered within the DGT, which might intervene if no effective competition existed (in this case, it was also required the favorable opinion of the competition agency) or if users complained of being abused. No subsidies (except for small public service obligations) or other guarantees were granted to overcome potential losses.

Concessionaires also retained an exclusivity right to operate services and infrastructures for 30 years in their lines (18, in short lines), including the right to build new ones within their right of way. However, to counteract this monopoly power and in order to promote effective competition among operators, concessions were designed to share several common tracks around major urban and industrial areas (particularly, Monterrey and Mexico City) and several ports (Tampico and Veracruz). For these cases, concession titles included detailed mandatory access and connecting rights between concessionaires. The prices of these rights were to be bilaterally negotiated between private operators, once they started operations, although the SCT should intervene if no agreement was reached before a year or when any of the concessionaires requested it.

As shown by Table 5, the first concession offered for sale, in June 1996, was the longest of the short lines, *Ferrocarril Chihuahua al Pacífico*, which the government thought could constitute a low risk test of its overall bidding system. Unfortunately, the railroad was in extremely poor conditions and only one bid for US$28 million was offered. Since this was below the government's reservation price of US$50 million, the sale was canceled in October 1996 and it was decided that the package should be restructured to attract more potential investors.

Table 5. Main results from concession sales completed until 1999

(*)	# Bidders (bid, in P$ billion)	Winning consortium	Transfer
Ferrocarril Chihuahua-Pacífico (SL)	• One bidder (0.02)	No bid above minimum	-
Ferrocarril del Noreste	• ICA/UP/SBC (4.1) • *Grupo Ferroviario Mexicano* (GFM) (4.2) • TMM/KCSI (11.0)	*TFM* (= TMM + KCSI)	Jun 1997
Ferrocarril Pacifico-Norte	• *Grupo Ferroviario Mexicano* (GFM) (4.1)	*FerroMex* (= GFM)	Feb 1998
U.F. Coahuila-Durango (SL)	• *Grupo Acerero Norte* (GAN) + Peñoles (0.2)	GAN/Peñoles	Mar 1998
U.F. Nacozari (SL)	• *Grupo Ferroviario Mexicano* (GFM)	No bid above minimum	-
Vía Corta Tijuana-Tecate (SL)	• Medios de Comunicación y Transporte (0.07)	Revoked due to no payment	-
Ferrocarril del Sureste	• GAN/Peñoles/Illinois (1.3) • TRIBASA (2.8)	*FerroSur* (=TRIBASA)	Jan 1999

Source: SCT (1996) and *Diario de la República Mexicana* (Official Gazette) *Note:* (*) SL= Short line.

In December 1996, the Northeast Railroad was acquired by *Transportación Ferroviaria Mexicana* (TFM), a consortium formed by a Mexican transportation company (*Transportación Marítima Mexicana*) and the US railroad *Kansas City Southern Industries*. With a bid of P$11 billion (US$1.4 billion) – almost three times the size of the runner-up's – TFM acquired 80% of the shares of the company: the first 32% had to be paid for soon after the auction, the next 48% within 180 days of the first payment, and the final 20% was planned to be acquired in 1999.

The second sale, announced in March 1997 was for *Ferrocarril Pacífico-Norte*, the most sought after of the three main lines. The concession documents allowed competitors to bid for the original Pacific-North concession alone or for a concession that also included the main connecting segment of the failed Chihuahua al Pacífico railroad (Ojinaga-Topolobamba). Although initially three consortia were interested, only one bid was finally submitted including the Ojinaga-Topolobamba line. The North-Pacific railroad was acquired by *FerroMex* in June 1997 for P$3.1 billion (US$524 million) for the 80% of capital and, as TFM had previously done, a 25% stake in the Mexico City's terminal company. The consortium was now integrated by the former losers *Grupo Ferroviario Mexicano* (74%), ICA (13%) and the US railroad, Union Pacific (13%), although ICA reached an agreement in December 1998 to sell its shares to Union Pacific. After private operations started in February 1998, *FerroMex* also acquired the 20% of shares remaining in government hands.

In October 1997 the short line Coahuila-Durango was concessioned for 30 years to a consortium integrated by Mexican firms *Grupo Acerero del Norte* (GAN) and *Industrias Peñoles*, two of the most important shippers, whose bid of P$180 billion was over the reservation price. The auction also included several other purchases and leases of rolling stock for about P$20 million.

The Southeast railroad, now *FerroSur*, was acquired in mid-December 1998 for US$322 million by the Mexican holding *Grupo Tribasa*, which also maintained interests in toll roads and airports. The winning bid for the 100% of the company was twice its only rival's, a consortium of GAN, *Industrias Peñoles* and *Illinois Central*. The main attraction of the Southeast railroad was the line connecting the port of Veracruz to Mexico City. *Grupo Tribasa* announced that it would not exercise the right to acquire the short line Chiapas-Mayab (comprising the railroads in the Yucatan peninsula), so this line was left to be privatized independently, along with the remaining short lines. The private operator took over operations in January 1999 and this transfer closed the sale of the three large companies into which the national network of railroads had been divided.

Finally, the Mexico City's Terminal, *Terminal Ferroviaria Valle de México*, is privately managed since April 1998. As scheduled, each one of the main rail operators owns 25% of the shares (included in their auction packages), whereas the government retains the remaining 25%.

At the end of 1999 a wide majority of private investors and government officials agreed that, particularly when compared to what had happened in the toll roads process,[7] railroad restructuring in Mexico constituted a fine example of transition from a model of public sector dominance to a system of private operation of an existing transport infrastructure.

---

[7] Mexico's toll road program suffered a backlash in the mid-1990s after the government had to rescue several concessions that had failed to find financial stability. For details, see for example Gomez-Ibañez and Meyer (1993).

4.3. Access issues in Mexican railroads

The restructuring process completely changed the dominant role played by FNM in the national railroad system since the 1940s and the way access issues had been formulated so far. At the end of 1999 FNM only operated a few short lines in the South (*Vía Corta del Sur*) and remained as the nominal owner of several short lines that had not been sold. Waiting for its liquidation, the company had leased most of these lines to the concessionaires of the main rail services, which received adequate compensation for public service obligations. With respect to passenger services, it was estimated that only 10 trains would be operating by December 1999 (down from 61 in 1996), requiring a public subsidy of P$164 million (US$17 million).

On the side of the private concessionaires it is still too early to carry out a detailed assessment of their impact on the sector's overall performance. However, the initial figures provided by the SCT seem to be positive. For example, it has been recorded that the new operators have already invested more than P$3 billion during 1997-1998 and it is estimated that another P$3.3 billion will be spent during 1999. According to the proposed business plan in their technical bids, the present value of investments during the first five years of private activity will be about P$9.0 billion.

Several factors can be identified in the railroad privatization model in Mexico related to competition and access among private railroads. The potential for this intramodal rivalry, which was one of the goals of the reform, is large but it could still be affected, positively and negatively, by three factors embedded in the system.

- **Intramodal competition favored by structural design.** When the packages of major rail lines to be concessioned and the short-lines related to them were designed, it was considered that, where possible, no concessionaire should have exclusive access to major cities (Mexico City, Monterrey and Guadalajara), industrial areas (Center-North of the country) or ports (Tampico and Veracruz) (see Table 4). This restriction required the mandatory imposition of trackage and haulage rights in the key routes, in order to grant a railway concessionaire access to other railway's licensed tracks, upon payment of a fixed fee. It also implied the limitation of exclusivity rights in the concession titles by not hindering other companies from operating the same routes, whenever they were willing to invest in parallel tracks.

  This design was particularly difficult in the case of the North-Pacific (*FerroMex*) and Northeast (TFM) networks, since the Southeast was connected to them only through Mexico City. These two railroads compete with each other in the Queretaro-Mexico City line, and in the access to Tampico, Aguascalientes and Monterrey. In the border crossings of Nuevo Laredo and Matamoros, TFM faced no competition, and neither did *FerroMex* over the crossings in Mexicali, Nogales, Ciudad Juarez and Piedras Negras. Both TFM and *FerroSur* had access to the largest port in the country, Veracruz, and all three concessionaires jointly operate Mexico City's terminal.

  Apparently, the effect of this intramodal competition design on the tariff levels has not been large yet. Although concessionaires must register their prices under the DGT and the SCT may intervene if "no effective competition" exists, no major complaints have been forwarded by the shippers at the moment. Prices have increased with respect to past years, but since services and quality have also done so, it is difficult to perceive a generalized negative response. Over the competitive tracks neither the authorized (maximum) tariffs nor the effective ones seem to differ too much among concessionaires, although detailed information on this topic is difficult to obtain. Since the definition of "effective competition" refers to lack of two or more rivals in the route, the risk of collusive practices could have been underestimated and its potential harm should be evaluated in the medium-term. With respect to the non-competitive routes, both the SCT and the competition agency retain a clear watchdog role and, since intermodal competition from the trucking industry is strong, no actions have been taken so far.

- **Conflicts on defining the access rights.** Although trackage and haulage rights were included in the concessions to favor competition among the operators, they could also pose several difficulties if they are not flexible enough. The 1995 railroad law (LRSF) ruled that the prices of these rights were to be bilaterally negotiated between private operators, although the SCT should intervene if no agreement was reached before a year or if the concessionaires requested it. This had not happened and in June 1999 a final agreement over this issue in the most conflictive case, between TFM and *FerroMex*, was reached.

  The huge difference in the bids made by each concessionaire and, particularly the lack of a detailed methodology on how to include these differences in the access prices was the major controversial issue that had prevented a previous agreement. The regulations developed by the LRSF were not very detailed and only requested the inclusion of the maintenance and operating costs, the incremental costs associated to the other firm's operation, depreciation and a reasonable profit for the provider of access. Since 1999 the DGT seems to be working on a detailed methodology to implement these prices if needed, although they could possibly lack enough detailed information to cope with this task. In the future, the problems could re-emerge, not only with the short lines and *FerroSur*, but also with the others, since the law also provided the possibility that concessionaires could negotiate additional trackage and haulage rights. In this case, the authorities could intervene only to review the agreements entered into.

  In general, the Mexican concessions cover a long period of time over which the transportation circumstances and economic environment may vary significantly. Thus, the transported cargo volume may in the future permit the coexistence of more than one carrier. Therefore, a more flexible mechanism for the assignment of trackage and haulage rights could be needed. Such mechanism should not discourage investment but rather allow the imposition of trackage and haulage rights whenever necessary and in the absence of effective competition.

  To overcome this possibility, the 1995 law allows the SCT to grant concessions to third parties, in order to provide transportation services (cargo or passengers) over a licensed track, but only after the specific exclusivity period (30 years for main lines, 18 for short ones) or whenever monopolistic practices have been engaged into by the concessionaire (previous opinion from the competition agency is required). In this case, the trackage or haulage rights which can be imposed do not include the right to serve intermediate points at the route subject to those rights and shall apply only for the transportation of a product or products for which feasible alternative transportation does not exist and for which the petition was made.

- **The operation of Mexico City's terminal.** A final source of potential access conflict among private operators is the ownership of *Terminal Ferroviaria Valle de Mexico* (TFVM), the concessionaire of Mexico City's complex 20-station network. The corporate governance of TFVM is rather peculiar, since it is jointly owned by the three main rail concessionaires (a 25% stake each). The remaining 25% (currently held by government) belongs to the future suburban rail operator. The owners are simultaneously the customers of TFVM, to whom they pay the services (not for access rights and slots, which are determined by a central traffic control). The firm apparently operates since April 1998 with total commercial autonomy and exquisite neutrality with respect to the owners.

  However, although this organizational form intended to prevent the external spread of conflicts, it also creates a long-run internal instability risk. A potential problem, for example, is the owners' asymmetry (in terms of traffic volume, number of connections with the Mexico City's network and even in the price paid for their concessions). Even so, they all have the same voting power and a majority of 75% is required for all decisions. If, for example, cargo volume discounts are introduced in the future, this could create fears of discrimination and trigger conflict.

  Finally, TFVM is now self-financing through its operational revenues. No additional equity was needed from owners apart from the initial disbursements and profits emerged just eight months after starting operations, thanks to cost control and improved performance. In the future, if additional capital if needed, the owners might seek to renegotiate their stakes.

5. DISCUSSION: WHAT ARE THE LESSONS TO BE LEARNED?

This paper has described the main common features of two important railway restructuring processes that had recently happened in Latin America, a region where – with few exceptions – the rail industry has gone further than elsewhere in terms of private participation. The Brazilian and Mexican cases share the common feature of representing a rail industry whose size is similar to any medium-size European country and that was completely dominated by a major government-owned firm before the restructuring. The spirit of the reform in these two countries – which also encompasses similar reforms in the region – is basically decentralization and disintegration of the public operator, seeking greater efficiency by increasing private sector participation. Our approach has followed this spirit, pushing the reforms one step further and drawing positive and negative lessons to be learned/avoided by European countries.

One important caveat, however, is the need to recall the differences between the European and Latin America transport systems. In the later case, freight (and particularly export-oriented cargo) constitutes the main traffic of lines that in most countries have suffered years of neglect and low maintenance. Moreover, in Latin America, the presence of the State in the economy is still large and the private sector capability is limited. In addition, legal limits often impose the concession formula instead of a more radical outright privatization mechanisms.

Taking into account these differences, one of the most important lessons learned from Latin America, with relevance and value to railway restructuring and reform in European countries. is that *concessioning works* when the government needs funds to pour in the transport sector without compromising too much other sources of public expenditure. The positive response of domestic and foreign companies in Brazil and Mexico demonstrates that even a narrow private capital sector can be quickly mobilized, creating a management expertise to provide both freight and passenger services under long-term concessioning contracts. In most cases, the risks associated with concessioning can be managed through bonded performance of the private firms, prudent diversification of concessionaires, and well-engineered contracts. However, the long-term viability on concessioning or any other privatization approach depends on competitive factors and the quality of management, which cannot be predicted in advance with certainty.

Therefore, by far the most important success factor in the process of concessioning is a *resolute political commitment* and clearly articulated objectives at the highest level of government. While the political and economic stakes are high throughout the privatization process, the greatest risks are at the front end when the program is conceived and articulated. Subsequently, it is essential that the government retain its commitment through unpopular (but essential) as well as through popular steps in the process. A country wishing to undertake a similar process must have the similar continuity in leadership and clarity in vision.

This is an important message for the European Union as well, whose rail policy advocates compromise prior to effectiveness. The Latin America experiences show that it is possible to address a competition-oriented transformation of the rail industry on issues such as organization, interconnection or access, in a relatively smooth way even though many countries considered at the beginning that it was not possible. It is obvious that international links and sovereignty issues are not so important in Latin America, but large countries in the region have showed that it is possible to re-organize complex transport systems in less than five years.

The third lesson is thus that any change can be addressed in an *open, contestable, simple, and easily understood way*. Unless it is effectively managed, the design of concessions, for example, can become a contentious and politicized aspect of the privatization process and can slow it down or even derail it. For instance, procedures for the evaluation of proposals should be well defined and clearly explained to all offers in advance of proposal preparation. A two step process of technical prequalification followed by “best and final” financial and technical proposals can be implemented more rapidly than a single round competition which is less defined in terms of expectations and offering terms. In any case, final evaluation criteria should be clearly defined, few in number, and quantifiable.

Therefore, advance preparation goes a long way toward determining a positive outcome. Bids are never better than the quality of the request for proposals to which they respond. The government should evoke realistic and workable proposals that can be translated into viable long-term contracts. Planning and evaluation criteria which reward optimism on the part of bidders may create a need to recompete the concession in a second round, or worse may cause optimistic assumptions be locked into non-viable contracts. Railway concessions are always difficult to value. Unclear or conflicting criteria may engender miscalculations on the part of bidders.

Another pragmatic lesson is that *getting it right is more important than getting it done.* As shown in Mexican case, if a first round bid is unrealistic, a second round may be needed or the government may need to sweeten the concession by assuming additional liability or be investing in concession prior to privatization. Concessioning is not necessarily a one-shot process. In addition, not all private sector ventures succeed, even under the best of circumstances. Hence, it is important to have a fallback plan for re-concessioning should the first attempt fail.

Finally, in particular reference to *access issues*, the main lesson to be learned is that, if possible, they should be *addressed during the concession design* stage. Both Mexico and Brazil opted for horizontal separation schemes that limited and controlled the number of interconnections among the concessionaires. They showed that access issues and charging principles can be setup in concession contracts, proving that – in principle – these complicated issues can be addressed through ex-ante negotiations. The role of the government could be then limited to arbitrage.

REFERENCES

Campos, J. and Alexander, I. (1999), "Post-Privatization Regulatory Problems: The Case of Brazilian Railroads", *The World Bank Transport Regulatory Framework Studies*, Washington, DC.

Campos, J. and Cantos, P. (2000), "Rail Transport Regulation", in G. de Rus and A. Estache (Eds.) *Privatization and Regulation of Transport Infrastructures: Guidelines for Policymakers and Regulators*. The World Bank Institute, Washington, DC.

Gomez-Ibañez, J.A. and J. Meyer (1993), *Going Private. The International Experience with Transport Privatization*. The Brookings Institution, Washington DC.

Gomez-Ibañez, J.A. (1997), "The Role of The World Bank in Mexico's Railway and Toll Road Systems", *mimeo.*, The World Bank, Washington, DC.

SCT (1996), "Investment Opportunities in the Mexican Railroad System", *mimeo.*, Secretaría de Comunicaciones y Transportes, Mexico DF.

The World Bank (1995), *World Development Report*. Washington DC.

Helsinki Workshop on

Infrastructure Charging on Railways

31 July - 1 August 2000

PROGRAMME

****

Date: Monday, 31 July
Location: Restaurant Ostrobotnia, Helsinki

****

9.00-9.30 *Coffee*

9.30-9.45 **Welcome**

9.45-12.00	**Session I:** Chair:	**Goals and Principles of Rail Pricing** Esko Niskanen
	Chris Nash (ITS, Leeds University)	Key Issues and Principles of Rail Infrastructure Pricing
	Gunnar Gustafsson (International Union of Railways, UIC)	Infrastructure Charges in Europe. A missed Chance to Increase Competitiveness?
	Tom Howes (European Commission)	Recent Policy Developments at the Commission

12.00-13.30 *Lunch*

13.30-15.45	**Session II:** Chair:	**Pricing and Competition** Bernard Caillaud
	Amihai Glazer (University of California, Irvine)	Differential Pricing and Mistake Avoidance
	Jerome Pouyet (CERAS-ENPC, Paris) (joint with Anna Bassanini)	Access Pricing for Interconnected Vertically Separated Industries
	Pedro Cantos Sánchez (Universidad de Valencia)	Vertical Relationships between Railway Operations and Infrastructure: an Economic Approach to the European Case

15.45-16.15	*Coffee*	

16.15-17.00	**Session III:**	**Pricing and Financing**
	Chair:	Stef Proost
	Bernard Caillaud (CERAS-ENCP, Paris) (joint with Jean Tirole)	Infrastructure Financing and Market Structure

17.00	*Bus to Hvitträsk*
18.00-18.30	*A tour of the Hvitträsk museum*
18.00-20.30	*Sauna and the lake*
20.30-22.00	*Dinner at Restaurant Hvitträsk*
22.00	*Bus to Helsinki*

****

Date: Tuesday, 1 August
Location: Restaurant Ostrobotnia, Helsinki

****

9.00-9.30	*Coffee*	

9.30-11.30	**Session IV:**	**Theory vs Experiences**
	Chair:	Amihai Glazer
	Stephen Gibson (Railtrack Plc, England)	Charging for the Use of Railway Capacity
	Heike Link (German Institute for Economic Research, DIW, Berlin)	Two-part Tariffs for Rail Access Charges in Germany
	Emile Quinet (CGPC-ENPC, Paris)	Path Allocation Issues from Theory to Practice. Application to France and Europe

11.30-13.00	*Lunch*

13.00-15.30	**Session V:** Chair:	**Experiences and Policy Conclusions** Chris Nash
	Tiina Idström (Finnish Rail Administration)	Estimation of Marginal Infrastructure Costs of the Finnish Rail Network
	Per-Ove Hesselborn (Swedish Institute for Transport and Communications Analysis, SIKA, Stockholm)	Swedish Experiences
	Katalin Tanczos (Budapest University of Technology and Economics) (joint with Gyula Farkas)	Railway Restructuring and the Preparation for the Introduction of Infrastructure Charging in Hungary
	Javier Campos (University of Las Palmas)	Rail Reform in Latin America: What Should Europe Learn/Avoid?

around 14.30 *Coffee*

15.30 **Summary**

Helsinki Workshop on

Infrastructure Charging on Railways

31 July - 1 August 2000

Participants:

Elina Berghäll, Government Institute for Economic Research (VATT), Finland

Bernard Caillaud, CERAS-ENCP, Paris, France

Javier Campos-Mendez, University of Las Palmas, Spain

Pedro Cantos Sánchez, Universidad de Valencia, Spain

Gunnar Eriksson, Ministry for Industry, Employment and Communications, Sweden

Stephen Gibson, Railtrack Plc, England

Amihai Glazer, University of California, Irvine, USA

Gunnar Gustafsson, International Union of Railways (UIC), Paris, France

Marja Heikkinen, Ministry of Transport and Communications, Finland

Per-Ove Hesselborn, Swedish Institute for Transport and Communications Analysis (SIKA), Sweden

Tom Howes, European Commission, Belgium

Tiina Idström, Finnish Rail Administration, Finland

Kaisa Kotakorpi, Government Institute for Economic Research (VATT), Finland

Heike Link, German Institute for Economic Research (DIW), Berlin, Germany

Markus Maibach, INFRAS, Switzerland

Mika Mäkilä, Finnish Rail Administration, Finland

Chris Nash, ITS, Leeds University, England

Anita Niskanen, Government Institute for Economic Research (VATT), Finland

Esko Niskanen, Government Institute for Economic Research (VATT), Finland

Jerome Pouyet, CERAS-ENPC, Paris, France

Stef Proost, CES, Katholieke Universiteit Leuven, Belgium

Emile Quinet, CGPC-ENPC, Paris, France

Tom Sansom, ITS, Leeds University, England

Erkki Siivonen, Government Institute for Economic Research (VATT), Finland

Tuomo Suvanto, Finnish Rail Administration, Finland

Katalin Tanczos, Budapest University of Technology and Economics, Hungary

Juha Tervonen, Electrowatt-Ekono Oy, Finland

Tellervo van der Lei, The Netherlands

Mall Villemi, Tallinn Technical University, Estonia