INSTITUT D'ECONOMIE INDUSTRIELLE

IDEI Report # 1

Rail Transport

January 2003

The Economics of Passenger Rail Transport
A Survey

IDEI Report #1 on
Passenger
Rail Transport

January 2003

The Economics of Passenger Rail Transport A Survey

The Economics of Passenger Rail Transport A Survey

prepared by

Paul Seabright

with contributions by

Claude Crampes
Etienne de Villemeur
Guido Friebel
Chris Glasson
Aldo Gonzalez
Marc Ivaldi
Jérôme Pouyet
Emile Quinet
Catherine Vibes

Acknowledgements

The authors acknowledge with gratitude the advice and research support of Deutsche Bahn AG, but note that the views expressed remain those of the authors alone and in no way represent the views of Deutsche Bahn AG or its representatives.

Contents

List of Tables

List of Figures

List of Boxes

1. INTRODUCTION

The objective of this document is to survey the main economic issues that arise in the management and regulation of passenger rail transport. The analytical framework we use is the standard Structure-Conduct-Performance paradigm created by the discipline of Industrial Economics. We will proceed in five main steps:

i) We begin with a definition of the product under scrutiny and a presentation of the characteristics of demand and supply of this product, drawing not just on general principles but on scientific research assessing the quantitative magnitudes of the various forces at work.
ii) The next step is to analyze how conditions of competition work in the industry. We study the various factors that determine how rail operators function: vertical and horizontal concentration, the diversification of activities, the differentiation of products and barriers to entry (and to exit).
iii) Next we consider the set of decision variables that matter for price and non price competition, in the short and in the long run,
iv) Next we discuss indices of private and public performance relevant to assessing how well rail transport is meeting the demands placed upon it.
v) Finally, we discuss regulation of the industry, including not just the rationale for regulation but evidence of the effectiveness of different regulatory regimes.

This work is to be viewed as a catalogue aimed at identifying the main problems that should be analyzed in depth in the future, taking into account interactions with other industries and the potential consequences of any change in technology, structure, strategy, or regulation[1].

[1] For SCP "activists" like Bain (1959) and Scherer (1980), there exists a *causal* stable relation between the structure of an industry, the conduct of firms and market performance. Here we limit the use of the SCP paradigm to the list of items of interest, without any causality presumption. Indeed, there is a powerful body of work that takes structure, conduct and performance to be jointly determined by underlying (mainly technological) variables.

2. THE CHARACTERISTICS OF PASSENGER RAIL SERVICES

2.1. Supply and demand – general principles

Are railways a unique industry, or can they be analyzed with the standard tools of Industrial Organization? For Campos and Cantos[2], "the rail industry poses a number of specific problems for transport economists and regulators that are only partially shared with other transport modes. These elements are the multi-product nature of the activity, the particular cost structure of railroad companies, the role of infrastructure and networks, the existence of indivisibilities in inputs and outputs, the organization of rail transport as a public service, and the existence of externalities in the transport system as a whole". Some of the above characteristics refer to the product (or service) itself, such as "multi-production", "indivisibilities" and "externalities". Others refer to the conditions under which the service is supplied. In the following, we try to distinguish the intrinsic character of the service from its supply and demand.

2.1.1. The basic characteristics of rail transport

Transportation as such consists in displacing an object with given characteristics from one point to another, starting and arriving at given dates under specified conditions of quality and safety. The first obvious characteristic of the transport product, therefore, is its very high differentiation, mainly in terms of geographical heterogeneity: since every trip can be defined by its departure and arrival nodes, in any country one can define an infinite number of different potential products. For railways, because most networks are meshed, it is also true that for each "product" there exists some flexibility in terms of its path. Concerning the time dimension, note that the product is hardly storable so that there exists little time substitutability even if there occur frequent delays.

Considering transport in such highly differentiated terms is not always helpful, however; sometimes aggregate categories of broadly similar services are more useful in practice. A standard distinction can be made between passengers and freight. For passengers, the main distinction is between long distance traffic and local traffic. In less developed countries, the former tends to shade into the latter, while in developed countries, the two types of products are clearly distinct. As regards freight, distinctions can be made in terms of dangerousness (implying the need for special wagons for inflammable and explosive materials), of the degree of care required (livestock *vs* inert material), and of speed (letters and newspapers *vs* bulk). Increasingly, too, there is a distinction between services that consist purely of transport between two rail depots, and those that consist of an integrated door-to-door service, with the potential for significant inter-modal cooperation.

Rail services tend to give rise to significant externalities, which are examined in more detail in section 2.4 below. Some of the most important and interesting of these are of the kind known as "network externalities", where the travel decisions of some individuals affect the costs or the benefits of other individuals elsewhere in the network and thereby the attractiveness of using the network as a whole. These can occur on the side of costs or on the side of demand, in sometimes complex ways we shall explore more fully below. For the time being we note merely that most rail services affect significant numbers of individuals in inter-related ways.

[2] See Campos and Cantos (1999).

2.1.2. The demand for rail transport

The demand for rail transport is essentially a derived demand – that is, one derived from the capacity of rail to satisfy a more fundamental demand for mobility. Usually each traveller has a demand for a specific service: from home to place of work and back, from home to vacation location and back, etc. Similarly, each parcel has a specific demand (though typically without a return); in what follows we shall consider mainly passenger demand. Such demand for a service translates into a demand for a mode of transport that supplies this service. Different modes can be thought of as supplying the service with different characteristics, which are substitutable for each other to varying degrees. The pioneering work of McFadden and collaborators since the 1970s has shown how we can estimate the extent to which users trade off the characteristics of different transport modes against each other. The insight they deploy is that different characteristics can be observed to influence the probability that a representative traveler will choose a given mode of transport, and the substitutability of these characteristics for each other is determined by the ratio of their effects on the choice probability. For instance, Table 1 reports regression results of a multinomial logit estimation of modal choice for work trips in the San Francisco Bay Area. To find the cost of time spent traveling in the vehicle as a proportion of the wage, for instance, the coefficient on the second regressor is divided by the coefficient on the first, yielding in this case that travelers on average find commuting time roughly half as costly as working time. Excess time (time waiting at the bus-stop, for instance) is considered more costly than working time, as is shown by dividing the coefficient on the third regressor by the coefficient on the first. This methodology has been extensively used to estimate elasticities of demand for rail transport with respect to changes in both price and travel time, and we survey the findings in section 2.2 below.

Table 1: Effect of cost and travel time on modal choice

Regressor	Parameter estimate	T-ratio
Cost/post-tax wage	-0.0412	-7.63
On-vehicle time	-0.0201	-2.78
Excess time	-0.0531	-7.54
Auto-alone dummy	-0.892	-3.38
Bus with auto access	-1.78	-7.52
Carpool	-2.15	-8.56

Notes: auto, carpool, bus with auto access; omitted variable is bus with walk access
Source: McFadden (1977).

The different characteristics of passenger services include speed, comfort, reliability and flexibility (or adaptability to the traveller's particular requirements). For instance, road congestion means that rail is often more reliable than road transport (in the sense that the arrival time is predictable to a greater degree of accuracy), while rail is less adaptable than road (in the sense that a car can take the passenger all the way to the final destination instead of only to a station along the way). Various intermediaries (such as travel agencies and logisticians) can increase the adaptability of rail transport by acting to bundle transport services (including rail services) more effectively than the individual customer acting alone.

The extent to which travelers value different characteristics will depend typically not just on their own characteristics as individuals, but also on the purpose of the travel. For instance, travel for business purposes is likely to involve a great deal more sensitivity to time and reliability than are most forms of leisure travel (this is confirmed by many authors, as summarized in Oum *et al*., 1990). It is not easy to obtain information on the division of transport demand by purpose of travel, but Table 2 provides such information for the United Kingdom based on survey data. Note that this table divides demand up according to the number of trips, without adjustment for the length or value of the trips concerned (rail trips will typically be longer than car trips, so that the share of rail in total trips is well below its share in the market as a whole). What the table shows is that the purposes for which people travel are many and various, and it is unlikely that passengers will react in a homogeneous fashion to changes in prices, service characteristics or network configurations.

Table 2: Trips per head of population per year by purpose 1998/2000 - United Kingdom

	Surface rail	All Modes	Share of trips by purpose/all purposes
Commuting	6	161	15.6%
Business	1	36	3.5%
Education	1	68	6.6%
Escort Education		48	4.7%
Shopping	1	216	21.0%
Other Escort		80	7.8%
Other Personal		103	10.0%
Social	2	242	23.5%
Holiday		29	2.8%
Other		46	4.5%
Total	11	1029	

Source: Department of Transport, Transport Statistics 2001 Edition.
Note: The number of rail trips in this table should not be interpreted as indicating a market share of rail, since rail trips are on average much longer and yield higher revenue than trips using other modes (which include those made on foot, by bicycle and so on).

Finally, it is important to note that demand for transport services has some of the characteristics of demand for network goods – in particular that the demand by one user is related to the number of other users also demanding the service. For instance, demand for travel to a particular destination is related to the number of other users also visiting that destination, not just because it is less costly to meet such a demand (which is strictly speaking a supply-side network externality), but because the number of other travelers itself changes the characteristics of the good concerned (for instance because a particular destination is known for its important business contacts or its attractive nightlife).

2.1.3. The supply of rail transport

There exists a large number of alternative means of transport. For a given service (initial node – final node) one can construct a merit order in terms of speed. But most vehicles incur a strong discontinuity at the origin because of technical or legal reasons (registration) or because of the topological characteristics of the networks on which they operate, as we show in Figure 1 below.

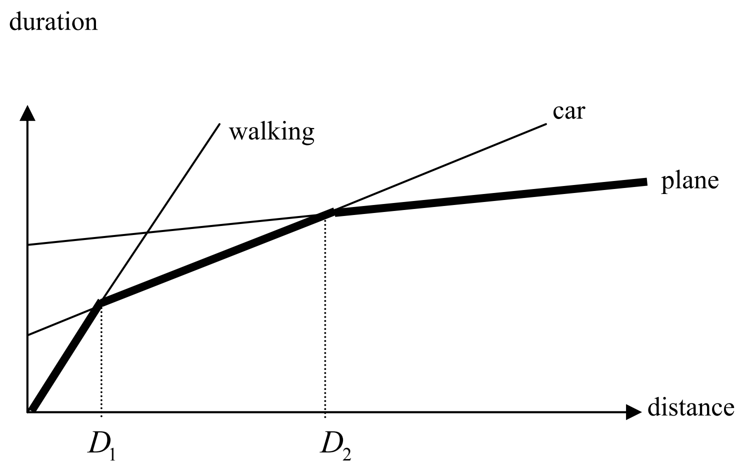


Figure 1: Time and Distance Travelled using Different Transport Modes

Rail transport can be a substitute for cars, buses or aircraft, or even for inland waterways depending on the distance and on the characteristics of the initial and final nodes. But although they may be to some extent comparable on the demand side (because they share desirable service characteristics to varying degrees), these different modes are very different in respect of both technical and economic properties. For instance, rail transport is highly capital intensive by comparison with most other modes. This arises from the high cost of acquisition, installation and maintenance of tracks and stations; as well as the cost of acquisition, operation and maintenance of the rolling stock. The result is that rail transport is characterized by strong economies of density, as well as some economies of network size. These economies of network size are the overall result of purely engineering considerations, which tend to favour scale, some economies of coordination, and some diseconomies of complexity in the management of large organizations (the overall balance between which can vary significantly between networks). There is also a strong need for coordination between infrastructure operation and service supply, though opinions have fluctuated over time as to whether such coordination requires vertical integration.

Historically, all firms have been multi-product suppliers, that is, suppliers of all the rail services. This makes it difficult to estimate costs of providing particular services because of the arbitrariness of allocation of the common cost components (as well as because, historically speaking, infrastructure and operations were often not even distinguished). In modern times, there has been some degree of specialization in the transport of passengers, for instance in the UK.

The function that describes the cost incurred by suppliers of rail transport is highly dependent on their degree of integration. For a train operator on a given line, the variable cost depends on the number of wagons[3] (as a proxy for the number of passengers or for the volume of freight), and on the speed[4]. These two variables determine the fuel and maintenance expenditures for the given distance. All the other costs are fixed, including the crew's wages and the company's administration costs, but also the depreciation of the rolling stock (when it comes from a leasing company), and the infrastructure operation and maintenance costs which are billed to the train operator. If the company operates on several lines, the variable and fixed costs change from one line to the other, which in

[3] Strictly speaking wagons are a fixed cost per journey, but on lines with fewer expected passengers journeys can be planned using fewer wagons. This example illustrates the familiar point that the difference between fixed and variable costs is not absolute but depends on the time frame under consideration: fixed costs are those that vary comparatively little with traffic within a given time frame. Wagons are more fixed than the catering costs, while track is more fixed than rolling stock.

[4] It also depends on how trains are organized: it is more costly in terms of fuel and crew to displace a given number of wagons with two locomotives (that is in two trains) than with only one. There are externalities between services too: removing one service typically means reschudling others.

continuous terms can be modeled as a function of distance. As regards infrastructure management, the cost is mainly a function of the distance and of the number of trains (tracks and signaling costs), of the number of passengers (stations) and of the volume of freight (terminal nodes). Infrastructure management also involves major logistical problems, notably ensuring that the rolling stock and train crews are available at the right time and place. These problems are often easier with larger networks, though not always, and are also highly dependent on the details of network structure. In addition there are organizational challenges due to:

- energy needs (connection to the power grid);
- the connection between final nodes (railway stations) and intermediary nodes (marshalling yards and maintenance stations);
- the operation of heterogeneous trains on the same line;
- reservation and billing systems;
- the organization of inter-modality (see section 3.4 below).

We summarize in section 2.3 below what is known empirically about the way in which costs are affected by different characteristics of the services and the operators that supply them.

2.2. What do we know empirically about demand?

Table 3 summarizes the results of studies estimating the elasticity of demand for passenger rail travel with respect to:

- the price of the rail services themselves;
- the price of competing modes of transport such as car and air;
- the time of travel by rail;
- the time of travel on competing modes.

The values found by most studies are surprisingly low, at least for short-run elasticities (nevertheless such findings are in line with the earlier survey of studies reported in Oum *et al*., 1990). This suggests that many passengers have strong preferences for the mode of travel they use, perhaps because they have invested in vehicles, jobs or (more generally) lifestyles that make substitution between transport modes difficult. However, estimated long-run elasticities are significantly higher, and over a longer time period it is clear that rail has been losing market share rapidly to other modes, though how much of this is because of a growing preference for flexibility, and how much to changing relative prices and travel times, is less clear. This is illustrated by Table 4

Table 3: Summary of Rail Elasticity Estimates from Studies in BTRE Database

Author	Year	Country	Price Elasticity									Time Elasticity								Type
			Own			Cross						Own		Cross						
			Short Run		Long Run	Air		Bus		Car				Air		Bus		Car		
Koppelman	1998	Canada	-1.40	-2.04								-1.55	-1.81							IC (1)
Mc carthy	1997	USA	-0.28	-0.61			0.04		0.01		0.07	-0.06	-0.61		0.01		0.00		0.05	IC
Taplin *et al.*	1997	Australia Sydney		-0.16					0.03		0.04									Comm
Wardman e.a.	1996	Britain		-0.59							0.25		-0.87						0.73	IC (2)
Brown *et al.*	1996	Ex URSS				0.13	0.36	0.21	0.31	0.17	0.35			0.13	0.34	0.17	0.33	0.14	0.35	IC
Gillen	1994		-0.22	-0.33																Urban
Industry comision	1993	Austr. UK Canada	-0.23	-0.62	-1.59							-0.16	-0.70						0.32	
Goodwin *et al.*	1992	Britain		-0.79																I.C.
Goodwin *et al*	1992	Britain		-0.65	-1.08															IC
Macket & Nash	1991	Britain										-0.58	-0.68			0.14	0.21	0.10	0.24	IC
Anrikopoulos e.a	1990	Canada		-0.22			0.15		0.01		0.06									IC
Luk & Hepburn	1993	Australia		-0.35																IC
De rus	1990	Spain	-0.18	-0.41																Urban
Ben-Akiva & Morikawa	1990	Netherland	-0.15	-1.50								-0.04	-0.34							
Chu	1989	USA Chicago		-0.37								-2.14	-3.00							Comm
Owen-Phillips	1987	Britain		-0.69	-1.08															IC
Pickarel	1987	USA		-1.18								-0.38	-0.97							IC
Lubulwa	1986	Australia		-0.43			1.12		0.40		0.02		-2.25		0.06		0.30		0.04	IC L.D
Morrison & Whinston	1985	USA	-0.57	-1.20																
Winston	1985	USA		-1.20									-1.58							IC
Winston	1985	USA		-0.86									-0.60							Urban

BTRE : Bureau of Transport and Regional Economics. Australia (2002).

Notes (1) Correspond to Toronto-Montreal Travel

(2) Leisure travel only

Table 4: Modal Shares in the Eu-15 1970-99 by passenger-kilometres

	Car	Buses & coaches	Air	Tram & Metro	Railways	Total
1970	73.9%	12.6%	1.5%	1.8%	10.1%	100%
1980	76.2%	11.6%	2.5%	1.4%	8.4%	100%
1990	79.1%	9.2%	3.9%	1.2%	6.7%	100%
1995	79.5%	8.7%	4.6%	1.1%	6.1%	100%
1999	79.0%	8.4%	5.4%	1.1%	6.1%	100%

Source: European Commission, D.G. Energy and Transport (2002).

Table 5 indicates that market shares of rail within the EU vary from country to country, from a high of 8.3% in Austria to a low of 1.4% in Greece. And there have been some strikingly different developments over the last decade, with market shares holding steady in Germany and collapsing in Portugal. Nevertheless, for most countries within the EU market shares are within the range of 4.5-8%. Table 6, by contrast, shows that passenger rail has obtained much smaller shares in the USA and very much larger shares in Japan: even in the late 1990s rail in Japan had a market share of nearly 30%, although that was significantly lower than the nearly 50% achieved in 1970. While it is clear that circumstances in Japan may be rather different from those in Europe, such figures do at least suggest that modal shares below 10% for rail are not fated to be a permanent feature of modern industrial societies.

Table 5: Railroad market shares by country

	1990	1995	1999
Austria	10.3%	10.4%	8.3%
Belgium	6.4%	5.9%	5.9%
Denmark	7.8%	6.7%	6.1%
Finland	4.9%	4.9%	4.7%
France	8.9%	7.2%	7.9%
Germany	7.3%	7.6%	7.7%
Greece	2.5%	1.6%	1.4%
Ireland	4.8%	4.0%	3.3%
Italy	6.7%	5.7%	5.0%
Luxembourg	3.7%	4.8%	4.5%
Netherland	6.7%	7.7%	7.5%
Portugal	9.1%	5.6%	4.0%
Spain	5.4%	4.6%	4.2%
Sweden	5.2%	5.8%	6.4%
United Kingdom	4.7%	4.2%	5.0%

Source: Union Internationale des Chemins de Fer (2002).

Table 6: Modal split in US and Japan

United States Passenger Transport
Mode split in %

	Car	Buses & coaches	Air	Tram & Metro	Railways	Total travel billion p-km
1970	91.3	2.3	5.4	0.5	0.5	3486
1980	89.1	2.4	7.7	0.4	0.4	4575
1990	86.6	3.2	9.5	0.3	0.3	6094
1998	85.5	3.3	10.6	0.3	0.3	7266

Source: US Department of Transport, Bureau of Transport Statistics (2002).

JAPAN Passenger Transport
Mode split in %

	Car	Buses & coaches	Air	Tram & Metro	Railways	Total travel billion p-km
1970	31.1	17.7	1.6		49.6	582
1980	41.5	14.2	3.8		40.5	776
1990	50.1	10.0	4.7		35.2	1102
1998	58.2	6.9	5.4		29.5	1341

Source: Transport Fact Book (2002).

Table 7 shows the growing share of high-speed rail travel in all rail travel in Europe. Here the message is that while high-speed rail has clearly tapped a source of demand it has not on its own provided the answer to reversing the decline in the market share of rail.

Table 7: High Speed Travel in Europe

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Belgium								0.6	0.8	0.8	0.9
Finland								0.1	0.1	0.1	0.1
France	14.9	17.9	19.0	18.9	20.5	21.4	24.8	27.6	30.6	32.2	34.7
Germany		2.0	5.2	7	8.2	8.7	8.9	10.1	10.2	11.6	13.9
Italy	0.3	0.4	0.4	0.5	0.8	1.1	1.3	2.4	3.6	4.5	5.1
Netherland									0.1	0.1	0.1
Spain			0.4	0.9	0.9	1.2	1.1	1.3	1.5	1.7	1.8
Sweden		0.1	0.2	0.3	0.3	0.5	1.1	1.3	1.6	1.8	2.1
Total	15.2	20.4	25.2	27.6	30.7	32.9	37.2	43.4	48.5	52.8	58.7
% of Rail travel	5.6%					12.3%				18.1%	

Note: Values in 1000 million Pass-km
Source: Union Internationale des Chemins de Fer (2002).

Overall, the principal messages coming from the empirical studies of demand seem to be twofold:

i) Demand for travel services is highly differentiated, and there is no single source of demand which is likely on its own to reverse the falling market share of rail in overall travel.
ii) Changes in relative prices, like changes in relative speeds, of rail and other transport modes, have very small short-run effects. It is only over a significant period of years that policies affecting these parameters are likely to have a large effect.

2.3. What do we know empirically about costs?

One of the most robust findings in the empirical literature concerns the nature of economies of scale. The literature distinguishes two measures of economies of scale: economies of density and economies of size. The first refers to the increase in cost with respect to the increase in traffic, keeping the size of the network constant. The second measures the change in cost with respect to the change in the size of the network, keeping density constant. This distinction is extremely important in transport studies, because in most transport modes there is evidence of significant economies of density. As the level of traffic rises on a given network (at least up to the point where the network becomes seriously congested), the cost per passenger-kilometre falls, since the greater traffic allows the costs of the network to be spread across the greater number of travellers. By contrast, the evidence is much less clear as to whether an increase in traffic lowers costs per passenger-kilometre when that increase comes about because of an expansion in overall network size.

Breautigam (1999) reports several studies accomplished in U.S. and Europe, in the period 1974 –1997, indicating that rail network presents significant economies of density but – on average - slight or non-existent economies of size. For the European Rail Operators, Shires and Preston (1999b) have computed the average value of both indicators for the period 1971 –1994 using a translog cost specification. Their results are in line with those reported by Braeutigam. Most of the firms show increasing returns to density but some evidence of decreasing returns to size, indicating that their networks may have been larger than was cost efficient. Both returns to density and returns to size were greater (not surprisingly) for small rather than large operators. They also calculate an optimal network length of around 3000 km, which if accurate would indicate that many of the European networks are oversized. However, such estimates are purely indicative and subject to large margins of error; they also take no account of technological change, or differences in network structure across operators (such as the difference between star networks as in France and polycentric networks as in Germany). They are also unable to take into account organizational improvements that may themselves influence optimal network size. They are also influenced by existing public service obligations which make many existing networks differ from optimal configuration, and they are subject to the statistical problem that size and density are themselves correlated. Nevertheless, they do at least suggest that simple “bigger is better” arguments find no support in the data.

In a more recent study, Cantos (2001) found that for the main European carriers returns to density range from 1.42 to 2.04 while returns to scale vary from 0.45 to 1.4, the average of the operators included in the sample being 1.73 and 0.83 respectively. In Table 8, ED refers to the estimated economies of density for each of the carriers studied. A value of 1 would indicate no returns to density – namely that a doubling of traffic on the network would leave costs per passenger unchanged. A value of 2 would indicate that a doubling of traffic on the network would halve costs per passenger (thereby showing that total costs were unchanged). Unsurprisingly most carriers have estimates of returns to density lying somewhere between 1 and 2. The standard error is a measure of the reliability of the estimate – roughly speaking, the true value has a 95% probability of lying within two standard errors on either side of the reported estimate.

Table 8: Returns to density and scale for European carriers

	ED	Standard Error	ES	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

Note: See Table 2: Economies of density and scale in Cantos (2001).
A star * indicates a result non-statistically different from 1 at 5%.
ED means Economies of density and ES means Economies of scale.

Some studies have addressed the issue of optimal network size for varying levels of traffic density. Preston (1996) found an optimal size of 4000 Km for traffic of 120 million train-km per annum. Mizutani (2001) performed a study for the Japanese urban private network finding estimates of traffic for 120 million train-km and an optimal size of 75 km (the fact that his study is based on an urban network appears to be what explains the dramatic difference between his findings and those of Preston, though the discrepancy also indicates the need for caution before relying too heavily on any particular estimates).

There are no studies looking at the effect of ownership and competitive conditions on costs in rail, in the manner of Ng and Seabright (2001) for air transport. It seems reasonable to suppose, though, that private ownership and the need to respond to commercial challenges would tend to reduce costs in a similar way to that found in airlines. This is in line with overall findings in the literature on privatisation in market economies, as summarised in Megginson and Netter (2001), who conclude that “research now supports the proposition that privately owned firms are more efficient and more profitable than otherwise comparable state-owned firms” (p. 380). Much depends of course on what is meant by “otherwise comparable” – in particular, there is not enough evidence for us to be able to say with confidence whether and to what extent commercialisation of state owned firms is able to act as a substitute for privatisation.

We consider the empirical evidence on the cost implications of different network configurations (integration versus non-integration, for example) in section 3.1 below. Here we conclude with the main messages of the literature on rail costs:

i) The unambiguous pattern of economies of density shown by all the empirical studies indicates that firms may decrease their average cost by making a more intensive use of their existent infrastructure, namely by attracting more traffic.
ii) Reductions in costs from increases in overall network size are of a much smaller order of magnitude, or even negligible.

2.4. Measuring externalities in rail transport

Externalities are an important feature of the economics of transport and especially of the rail industry. Much effort has been devoted to estimating these externalities and especially to translate them into monetary values. Here we aim to synthesize the research and studies in this field.

We begin by recalling some economic definitions. An externality is defined as an effect by which an agent is affected by another agent, the effect being not channelled through the market. Examples include:

- I make a noise and disturb my neighbour.
- I play a Beethoven symphony on my CD set and my neighbour enjoys it.
- I enjoy freely – or to a price below cost - the use of a motorway. It is a positive externality for me, and a negative one for the tax-payers who pay the difference.

These examples allow us to draw a typology:

- There are positive externalities (second case) and negative externalities (first case).
- There are technical externalities (the two first cases, due to a "physical" fact) and pecuniary externalities (the third case), related to transfers and imperfect pricing.

Box 1: Technological Externalities and Pecuniary Externalities

Technological externalities lie outside the price mechanism; they arise when one agent or firm directly affects the utility of another agent or the profits of another firm by its actions; for example, by depositing waste material in a river a firm reduces the quality of fish in the river and reduces the satisfaction of fishermen, or raises the costs of production of another firm by making it purify the water which it needs to use.

Pecuniary externalities correspond to the idea that one agent benefits (or suffers) from the action of another without paying the cost (or being compensated), bust where the cost or the benefit is reflected in the price. For example, the hairdresser in a large urban area may see her wages increase, not because her productivity has increased, but because other wages have increased; or the rent for a given property may rise and benefit its owner due to the provision of some public services in the area.

Rail services are the source of many externalities, the list of which is somewhat arbitrary. We take into account the following ones:

- Pecuniary externalities due to imperfect pricing;
- Public service obligations;
- Environmental externalities;
- Congestion and safety;
- Impact on economic development.

We shall not dwell much on the first two points: pecuniary externalities due to imperfect pricing, and public service obligations. They depend on legal and institutional arrangements and are closely linked to history and political choices. Furthermore, they raise no particular conceptual problems, though they raise many statistical ones.

2.4.1. Pecuniary externalities due to imperfect pricing

Imperfect pricing is a kind of pecuniary externality. The main imperfection in transport pricing is due to infrastructure charges. The doctrine of the European Commission is that, in order to reach efficiency, charges should be set equal to the short run marginal social cost, including:

- the marginal cost of infrastructure damage;
- the marginal external cost of congestion;
- the marginal external cost of pollution;
- the marginal external cost of accidents.

Note that these do *not* include the fixed costs of infrastructure provision. Economic principles would suggest that, to the extent that infrastructure is congested, long-run marginal cost pricing is the appropriate concept. At all events, calculation of externalities suggests that actual infrastructure charges are rarely in line with even short-run marginal costs, sometimes lying above but more often significantly below.

2.4.2. Public service obligations

A second kind of pecuniary externality is due to public service obligations. Under this category can be listed:

- The obligation to keep extra staff for employment reasons;
- Retirement conditions;
- The construction of unprofitable tracks;
- The operation of unprofitable services;
- Low tariffs for some categories of users (unemployed, elderly…)[5].

Some of these PSOs are subsidized (in France items 2, partly 3 and 5), while others are hidden and not subsidized. The calculation of subsidies is very controversial and manipulable.

2.4.3. Environmental externalities

The impacts of transport on the environment are as many and various as the idea of the environment is vast and elastic. We have adopted a fairly wide perspective, including in the analysis the effects listed in Table 9.

[5] By this must be understood lower tariffs than those that would be set in any case by a profit-maximising firm, that would find it profitable to set different tariffs for different categories of user.

Table 9: A taxonomy of the impacts according to their sources and their targets

Type of impact	Targets			
	Landscapes	Fauna, flora, ecosystems	Material goods	Human beings
Aesthetic effects	Mainly visual effect of existence of infrastructure			
Impact on ecosystems		Infrastructure impact on soil, underground water & habitats		
Noise and vibration		Health damage to animals	Damage to buildings	Damages to health (stress, cardio-vascular disease) and loss of amenity
Local and regional air pollution	Atmosphere appearence	Damage to crops and fauna (acid rains) due to Nox, Sox, VOC and particulates from motor exhausts	Damages to buildings due to dusts deposit and chemical effects	Damage to human health due to Nox, Sox, VOC cardio-vascuclar diseases, pulmonary diseases) and particulates (cancer) from motor exhausts
Global air pollution	Green House Effects (mainly CO^2) leading to higher sea level	Changes in faunas and flora biotopes due to increase in temperature		Changes in the level of wealth and income, various from a country to another
Spatial Impacts				Changes in location (eg urban sprawl due to car use)
Non-renewable natural resources			Decreased stocks of some non-renewable resources: mainly petrol and some water	

2.4.4. Ascribing monetary values to externalities

In order to inform economic decisions, the measurement of externalities needs to go beyond a taxonomy of possible effects. There are several levels of information required. The first is an accurate description of the phenomenon. For certain impacts (such as effects on the landscape) it would seem difficult to go beyond this, though we shall see later that techniques exist that can help us to do so.

Where it is possible to quantify, we obtain greater comparability and thus better guidance for decision-making. A further step is taken by creating a hierarchy of impacts, and the most comprehensive way to do this is to attribute monetary values. It is clear that quantification, and furthermore any weighting of a monetary kind, imply a subjective hierarchy, different from one person to another, of the relative importance of these various effects.

The most important techniques for ascribing monetary values to environmental effects (see Quinet, 1994) are:

Surrogate markets, notably:

- The cost of trips necessary to benefit from some amenities can be used to value these amenities; this method is used, for example, for the evaluation of leisure parks.
- The method of hedonic prices: the price of some marketed goods depends on the characteristics such as air pollution or noise exposure. Variations in the prices of these goods with respect to variations in the environmental factor can thus be used to estimate the implicit value that individuals attach to the environmental quality.
- The estimation of the cost of environmental protection. Observing the amount which individuals are prepared to pay to reduce or eliminate negative environmental effects provides an estimate of their implicit valuation.

These methods present several difficulties. For instance, hedonic relations are not true demand functions: the residents of the most exposed houses may have chosen them partly because they are less sensitive to the environmental factor, e.g. noise, in which case the value of dwellings would decrease less than if the sensitivity were the same for everybody. Furthermore, there is the problem of distinguishing the individual impacts of each of the environmental variables, which are often closely linked to each other. Moreover, we have to assume for accurate hedonic valuations that people are well informed and aware of the damage caused by the nuisance at stake.

Contingent valuations

These consist in asking people what they are willing to pay to avoid the nuisance, or what they are willing to receive in order to continue suffering from it. Difficulties of implementation are, however, numerous:

- In order to get reliable answers, it is necessary to devote a lot of care to devising the questionnaire, and to administer it through very sophisticated procedures.
- Psychological biases lead to the result that willingness to pay to avoid a nuisance is consistently lower than the willingness to accept compensation to continue to suffer it.
- As with other methods, the results can be biased by poor knowledge of the actual damage.

Indirect methods, notably:

- estimation of avoidance cost;
- estimation of damage cost.

These are implemented in two stages. The first is technical and aims at estimating the consequences of the nuisance in physical units; for instance, in the case of air pollution, the frequency and importance of any impacts on health, or the damage to buildings. The second stage is the monetisation of these damages, either through market prices for the damaged goods, or through the cost of repairing the damage, the health care costs of injured or sick people, or through more subjective valuations (e.g. the value of human life). This method gets round the problem of the lack of information, but monetary valuation of environmental damage is hazardous. For example, the cost of repair overestimates the willingness to pay when the repair is not actually carried out, and the repair implies the definition of a zero level for the nuisance, and this definition is often arbitrary (in the case of noise for instance).

As can easily be seen, the methods of surrogate markets and contingent valuations estimate the willingness to pay of those who bear the nuisance; they deal with demand. Avoidance or repair costs deal with supply. If the decisions were optimal, these methods would produce the same result and the double equality would hold:

Price = marginal cost = marginal willingness to pay

However, in practice these methods do not provide equal values. Thus we have to take the decision as to which one to choose. The answer will typically depend on the nature of the problem to be solved (Quinet, 1994, Mauch and Rottengather, 1996).

Moreover, both cost and willingness to pay methods imply several different actors. In the case of noise, the willingness to pay comprises the willingness to pay of any of the agents who suffer from the noise. Similarly, the function of protection costs includes both the cost of a noise barrier, paid for by the public authorities, and the costs of double glazed windows paid by the residents.

When the effect is permanent (in the case of the global warming for instance), its valuation needs to include the effects on future generations, a point which raises two problems: how to estimate the value for these future generations and what discount rate to choose. It is often argued that three categories of values need to be taken into account, the sum of which is the total value of the good:

- the value in use, which corresponds to the actual revealed consumption of the good;
- the option value, which corresponds to the possibility of a future consumption of the good;
- the existence value: which corresponds to the value obtained by an individual who never consumes the good, appreciates its existence so that other people can use it, or simply just as a resource.

These three values also exist for marketed goods, but it is often thought that option value and existence value are pre-eminent in the case of environmental goods, especially for the case of depletable goods. Cost of travel or hedonic prices provide use values only; for contingent valuations the point depends on how the questions are asked.

We now consider how these techniques are applied in the analysis of rail transport externalities.

2.4.5. Time externalities and their various forms: the "congestion" complex

Congestion is an imprecise term, which covers two different phenomena. The first occurs in operations, when the delay of a service induces delays in adjacent services that are close to the first; this effect will be termed pure congestion in what follows. The second takes place when a service cannot be scheduled at a proper place, and induces a gap between the desired and the effective arrival time. We first present the situation for road and air transport, then for rail.

Transport users spend time on making journeys, which can be expressed in money terms, through the concept of value of time. The externality stems from the fact that travel time is not fixed, but depends on traffic levels. This point is often expressed by saying that transport is a variable quality good (Levy-Lambert, 1968) or subject to congestion (Kolm, 1968). This variability varies according to the mode. It is important, however, to appreciate that congestion operates in quite different ways in the different modes of travel.

Road congestion

Road congestion depends on the laws of traffic flow. Limiting ourselves to macroscopic analysis (as opposed to microscopic analysis which tries to model the behaviour of each vehicle in the flow), and taking into account just the situation of the flow of a given number of vehicles along an infinite road, the following relation holds:

$$D = \frac{V}{Q}$$

where:

- speed (distance covered by the vehicles in one unit of time): *V;*
- flow (number of vehicles passing through a given section of the road within one unit of time): *Q;*
- density (number of vehicles on a unit length of road at a point in time): *D.*

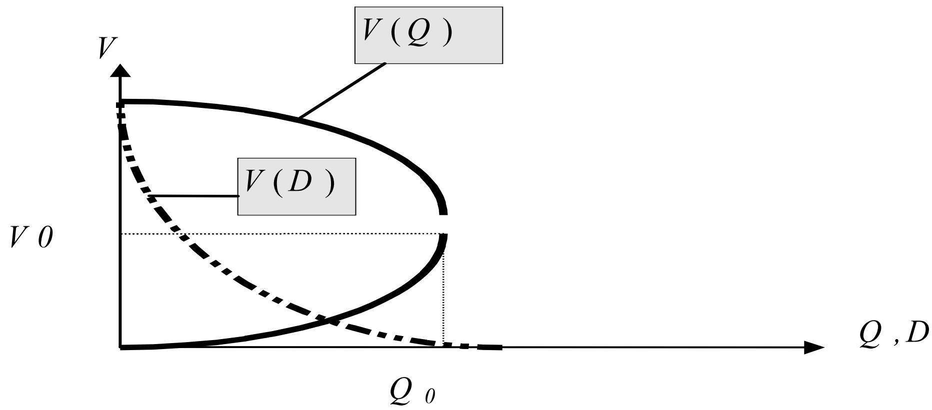


Figure 2: Relation between the flow Q and the speed V

Experience shows that speed *V* decreases with density *D*. It follows that the flow *Q* is related to speed according to the curve *V(Q)* which has the shape indicated in Figure 2. The upper part of this curve corresponds to a stable flow, the lower part to an unstable flow, where there is a low average speed and a lot of stop and go, which happens in a situation of high density. Q_0 is the capacity of the road.

Air congestion

Air congestion is quite different from road congestion in that it is not simply expressed through an increase in the travel time. Its main effect is a gap between the desired and the effective arrival time. When total traffic exceeds the capacity of the runways or of the air control system, flights have to be postponed or brought forward; the scarce resource constituted by the slots can be distributed arbitrarily by the public authorities, or allocated through procedures such as auctions, or through "grandfather rights" procedures which give priority to the existing users. Congestion is planned, it is not directly visible to the user, except in terms of an inconvenient timetable, and flight times are not increased as long as there is no unexpected incident.

There is another type of air congestion, when an incident delays a flight, and the immediately following flights have to be delayed in consequence.

Rail congestion

In rail travel, scarcity is manifested in the gap between desired and effective arrival time. But the speeds of all trains are not the same and their stopping patterns can differ. Unlike planes, faster trains can only overtake slower trains according to a pre-determined pattern according to the configuration of the tracks. This point clearly appears on a graph of time paths, as shown in Figure 3 which presents the example of trains between station A and station B. Each train is represented by broken lines, the slopes of which are proportional to the speed of the train on each section, the stops are represented by the steps in the line. This graph shows several features of rail scarcity.

When trains have the same speed (as in the case of trains 3, 4 and 5), capacity is high. A slow train requires much more capacity (case of train 6). In order to allow fast trains not to be penalized, it is necessary to stop a slower train in order to allow the faster train to overtake (case of train 1). The characteristics of rail scarcity are then clear: it does not occur with intercity links where dedicated tracks exist as in the case of HST, or where passenger intercity trains are run during daytime and freight trains are run during night. But congestion can be serious around large cities, especially at peak hours when the traffic is composed of a mix of trains with different speeds: slow commuting trains, fast intercity trains, and slower freight trains (especially for combined transport), all having strict arrival times, or on intercity routes where different trains have different stopping patterns to meet different traffic objectives.

As for air traffic, congestion may arise when an incident on a train delays the following trains.

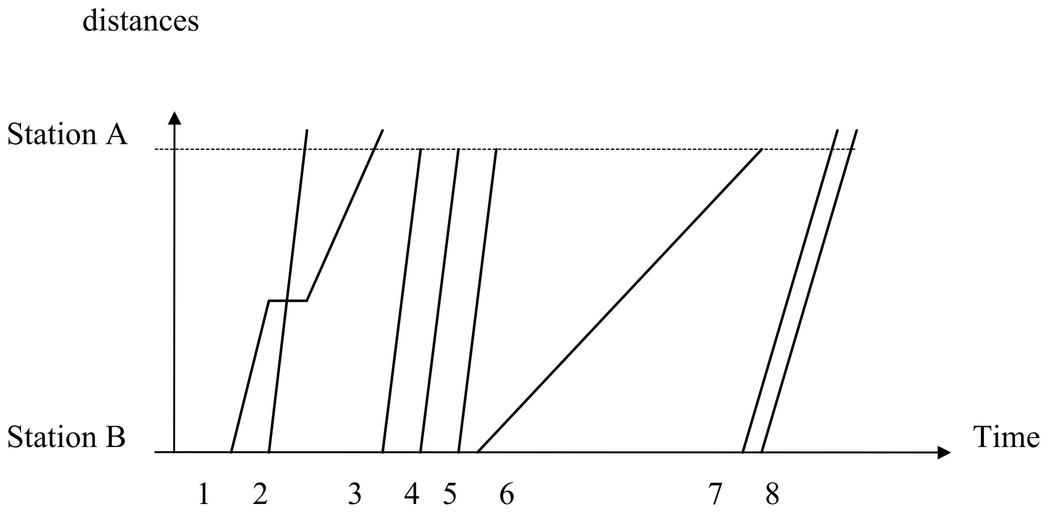


Figure 3: Railway Paths

Box 2: The Mohring effect (drawn from Small (1992))

Assume an urban public transport line, the flow on which is q; the operating cost of the buses is C_v, each having a capacity of *N*, and the time interval between them is e. The travel time is assumed to be θ, and the value of time is unity.

The operator has to solve the following program :

$$\underset{e}{Min}\, C = q\theta + q\frac{e}{2} + \frac{c_V}{e} = C^* \qquad \text{with:} \quad q_e \leq N$$

There are two cases to consider:

1) $q^e > N$

The optimal interval between two buses is $e = \sqrt{\dfrac{2c_V}{q}}$

The cost for the operator is:

$$\frac{C_v}{e} = \sqrt{\frac{qC_v}{2}}$$

which is an increasing returns to scale cost function, and the total social cost is:

$$C^* = q\theta + \sqrt{2qC_v}$$

2) $q_e = N$

the total social cost is: $C^* = q\theta + \dfrac{N}{2} + q\dfrac{C_v}{N}$

The cost for the operator is :

$$q\frac{C_V}{N}$$

which is a decreasing returns to scale cost function:
The move from the first process to the second one occurs for a flow such that:

$$\frac{N}{q} = \sqrt{\frac{2C_V}{q}}$$

or

$$q = \frac{N^2}{2C_v}$$

In this limit situation, the waiting cost is equal to the operation cost.

Table 10: Values of Time (VOT) Based on Recent Studies

Relevant VOT studies	HCG 1994	HCG 1998	HCG 1998	SNRA 1997	EUNET 1998	UNITE Values
Transport Segment			Euro 1998			Euro 1998
Inflation to 1998 Transfer to Euro						Normal Travel
Passenger transport – VOT per person-hour						
Car/motorcycle		6.70		9.31		
Business	21.33	21.00		11.95		21.00
Commuting/private	5.53	6.37		3.91		6.00
Leisure/holiday	3.79	5.08		3.10		4.00
Coach (Inter-urban)						
Business	21.23					21.00
Commuting/private	5.95			5.40		6.00
Leisure/holiday	3.08			4.37		4.00
Urban bus/tramway						
Business	21.23					21.00
Commuting/private	5.95			4.94		6.00
Leisure/holiday	3.08			3.22		3.20
Inter-urban rail		4.97		8.50		
Business		18.43		11.95		21.00
Commuting/private		6.48		6.21		6.40
Leisure/holiday		4.41		4.94		4.70
Air traffic					40.60	
Business				16.20		28.50
Commuting/private				10.11		10.00
Leisure/holiday				10.11		10.00
Freight VOT						
Road transport						
LGV	39.68	30.75	40.76			40.00
HGV	39.68	30.75	43.47			43.00
Rail transport						
Full trainload		645.37	725.45			725.00
Wagon load		26.16	28.98			30.00
Average per tonne			0.76			0.76
Inland navigation						
Full ship load		178.55	201.06			200.00
Average per tonne			0.18			0.18
Maritime shipping						
Full ship load		178.55	201.06			200.00
Average per tonne			0.18			0.18
Air transport						
Average per tonne						4.00

Source:
- Hague Consulting Group (1994), *UK Value Time Study*.
- Hague Consulting Group (1998), *The 1985-1996 dutch VOT Studies*. Paper presented at the PTRC Seminar on the Value of Time, Wokingham, UK.
- EUNET 1998: *Socio-Economic and Spatial Impacts of Transport Measurement and Valuation of the Impacts of Transport Initiatives* (John Nellthorp, Peter Mackie, Abigail Bristow, ITS Leeds)
- *Swedish National Road Administration*, guidelines for project appraisal (1997).

The Mohring effect

In public transport, there is another external effect stemming from the increase of traffic. This effect lead to a positive externality, not to a negative one: when traffic increases, the operator is led to increase the frequency of the services, thus by reducing the waiting time total travel time of the users.

There is an externality, but it is a positive one. Box 2 models this phenomenon according to a process analyzed in Small (1992).

Values of time

The monetization of congestion uses values of time, which act as multipliers of time losses expressed in hours. Many studies have been undertaken to measure values of time. Table 10 shows the results of various such studies, including the most recent comprehensive study of infrastructure charging and transport accounting, the UNITE program. The main message to come out of the table is the significant difference between the value of time for business travel and that for leisure travel.

Accident costs

The valuation of the cost of accidents is the multiple of the number of fatalities and injuries by the monetary value placed on a human life or on an injury. The value of human life is a statistical concept, representing the value of an unidentified life, not one known to the decision-maker. It also represents a marginal value, corresponding to the valuation of a small risk to the life of that person.

Two methods are currently used. The human capital method is the oldest, based on the calculation of the discounted loss of production caused by the death. The loss of production can be calculated on either a net basis (excluding the discounted consumption which will not take place) or a gross basis, not excluding this consumption. The implementation of the calculation faces several difficulties: the choice of discount rate, to which the result is very sensitive; uncertainty over future levels of production, over a long period; and ethical considerations, since the values associated with several groups of people, pensioners or the unemployed, are zero. In addition, this method does not take into account the value of pain and suffering.

The second method is based on willingness to pay, which can be implemented through two different procedures. The first is based on revealed preference, such as the extra wage for risky jobs; the second is based on contingent valuation, through interview techniques. Contingent valuation is open to the difficulties previously identified, plus the problem that people are not aware of the actual risks from transport, and are not able to identify and evaluate the consequences arising from the very small risk levels implied by safety in transport. This is clear form the differences in the two possible values: what individuals are prepared to pay to reduce an already very small personal risk of death or what they are prepared to pay to reduce the risk of a fatal accident in the community at large.

Results are again widespread, the values given by human capital methodology being generally lower. Here are the values of statistical life used in the European program of Transport Accounting Methodology UNITE, these values being drawn mainly from contingent valuations.

Table 11: Proposed UNITE Values Of Statistical Life (VOSL) by Country and compared to official values

Country	Official values in use A) million €, 1998	UNITE VOSL million €	(Official-UNITE)/Official Percentage
Austria	1.52	1.68	10
Belgium	0.40	1.67	312
Denmark(B)	0.52*)	1.79	244
Finland	0.89*)	1.54	73
France	0.62	1.49	141
Germany	0.87	1.62	87
Greece (B)	0.14	1.00	588
Ireland	1.04	1.63	57
Italy	n.a.	1.51	-
Luxembourg	n.a.	2.64	-
Netherlands	0.12	1.70	1269
Norway (B)	1.49	1.93	29
Portugal	0.04	1.12	2896
Spain	0.07	1.21	1625
Sweden (B)	1.48*)	1.53	4
Switzerland (B)	n.a.	1.91	-
United Kingdom (B)	1.53*)	1.52	1
Hungary (B)	n.a.	0.74	-
Estonia (B)	n.a.	0.65	-

Note: A) Based on Nellthorp, Mackie and Bristow (1998). HICPs (Harmonized Indexes for Consumers Prices) issued by Eurostat for Eurozone has been used to adjust price level to 1998.
*) Latest available values from Tervonen (1999) has been used (For Sweden SIKA(2000)). Corresponding EUNET values are; DK 0.79; FIN 1.33; N n.a. ; S 1.80; UK 1.11.
B) Not in Eurozone, Exchange rate of 24 November 2000 used.

A further question is whether altruism influences the value derived from willingness to pay methodologies. Jones-Lee (1989) suggests that the value of human life procured by this method should not be modified in order to take altruism into account, provided that altruism is not influenced by the kind of good at stake (i.e. there is no paternalism).

Another question is whether the value of human life should be the same in all transport modes. Let us explore this point, taking as a basis the value of human life in road transport, and considering the reasons why there should be any difference with other modes. The first, and simplest, reason is the difference in the socio-economic characteristics of the users of each mode. Of these, income is probably the most important factor; for this reason, the value of human life should be lower in urban public transport than for car users, and higher in interurban public transport, less so for standard or coach class passengers on trains or planes, more so for first or business class users. Note that this does not imply anything about the “absolute” or objective value of the lives of different individuals: instead, it is a reflection of the fact that some people are more willing than others to choose safer modes, in much the same way as those with a higher value of time will choose faster modes. Though this point is logically sound, it has never been introduced in transport modelling, which means at least that its size has not been important in estimates. Another reason why this effect does not appear is that people are not aware of the real conditions of safety in transport. Viscusi (1993) has shown that people underestimate risks in private transport, and overestimate them in public transport.

Another reason for a discrepancy between road and other modes as far as value of life is concerned is the correlation between the effort to reduce the risk and the willingness to pay for safety. According to Kolm (1968), an improvement of a road implies a reduction of risk (for example through more prudent driving behaviour) which is greater for those who have a high value of life than for the others who trade off the possibility of increased safety which is offered to them against the possibility of higher speed on the improved road. The consequence is that public decisions should be taken with an average value of life lower than the average willingness to pay for safety of the users.

Jones-Lee (1989) has suggested two more reasons for the difference between road and other modes: first the context effect, which means, for example, that people have a higher cost dying in a tunnel than dying in the open air. The difference between modes could be up to 50%. A second cause is the mass effect: when a large number of people die together, the value of life of each of the fatalities is higher. This point is made clear by the impact on the media, but contingent analysis surveys of this issue do not show any clear statistical result.

These reasons support the idea of a value of life which is higher in public transport than in road transport. Jones-Lee and Loomes (1994) suggest a ratio of 1,5 to 2. Two main considerations seem to be at work here. First, as public transport users have no direct responsibility for safety, they need to be more protected (it is possible to justify this position through a model of imperfect information of the operator about the value of life of the users, coupled with decreasing returns in the technical function of safety procurement). Secondly, the mechanism of legal responsibility induces the managers of a public transport operator to spend more money than would an agent working with his own money.

Other accident costs

In order to calculate the costs of accidents, it is necessary to take into account, not only the value of human life, but also the cost of injuries and damage to property. The valuation of damage to property is of course based on the monetary cost of the damage. The value of injuries adds direct costs (medical care, transport of the injured people…), indirect costs (loss of production) and an estimation, very subjective, of the pain and grief.

Impact on economic development

The classical economic analysis of the effects of transport on economic development, as used for instance in project appraisal, makes the assumption of perfect market competition, a condition which implies, inter alia, that:

- there are constant returns to scale and no monopoly;
- there are no external effects;
- space as such is unimportant.

These assumptions evidently do not hold in practice:

First, the existence of agglomerations implies that some forces are at work which fight against dispersion, the type of organization which would be predominant if all activities experienced constant returns to scale: then an even repartition of activities would minimize the total production and transport costs.

Secondly, historical evidence indicates that growth rates of poor regions show no systematic tendency to converge to those of rich regions, as would occur if there were constant returns to scale.

Several explanations have been advanced to explain these discrepancies:

i) The role of public capital and especially of transport facilities in the process of economic growth has been emphasized since the 90s (in stream of work initiated by Aschauer). In this line of thought, not only labor and private capital are factors of growth, but also public capital, for instance transport infrastructure capital. The magnitude of this factor has been assessed through statistical analysis of production – or cost - functions at the macro-economic level or at the level of economic sectors. The results are very scattered and show that there is a positive effect of public capital, but that it is difficult to assess its magnitude; the first studies showed that the effect was very high, but more refined studies tend to show that in fact the effect is rather low. A consequence of this effect would be that the social rate of return of infrastructure investments as reckoned by the usual methods would be underestimated, but there are doubts on the level of this underestimation, and no country has introduced this factor in its official manuals of project appraisal.

ii) Increasing returns to scale are more frequent than assumed in the classical economic text-books. For instance, increasing returns to scale exist in specialized services to industrial firms or sophisticated consumption goods.

iii) Space induces many peculiarities in economics, and it is often thought that spatial interactions are one of the main possible causes of public capital productivity. For instance, proximity may induce communication externalities (better spill-over of productivity improvement), or agglomeration externalities (better matching of demand and supply, an increase in diversity of available goods or production factors).

There are many possible mechanisms and models associating these features (public capital productivity, increasing return to scale, positive externalities). The main results of theoretical models implementing these mechanisms are that, usually, a decrease of transport costs induces:

- an increase in market areas and in competition between firms (efficiency effects);
- changes in locations of activities (distributional effects), marked either by agglomeration and polarisation in order to benefit from larger markets and more varied services, or less frequently by dispersion in order to avoid competition;
- location effects can be observed around new infrastructure through ex ante-ex post analysis. These analyses have been systematically done in France for High Speed Trains and Motorways, and provide conclusions that are confirmed by observations in other countries;
- motorways have effects on tourism and service activities. There is an increase of economic activity around the lay-out (about 20 km), and an increase of market areas of firms. Urbanization experiences developments around the interchanges;
- TGVs have effects on high level services, provoke changes in the internal organization of firms, and urban development located around the stations.

It is not possible to provide a magnitude of these effects, even less possible to distinguish between the effects of the various kinds of public capital, even less between the effects of the various transport mode infrastructures. Experts speak of a rough magnitude of 10 to 30% of the level of direct effects of transport improvements, but they are not unanimous on these figures.

Numerical results

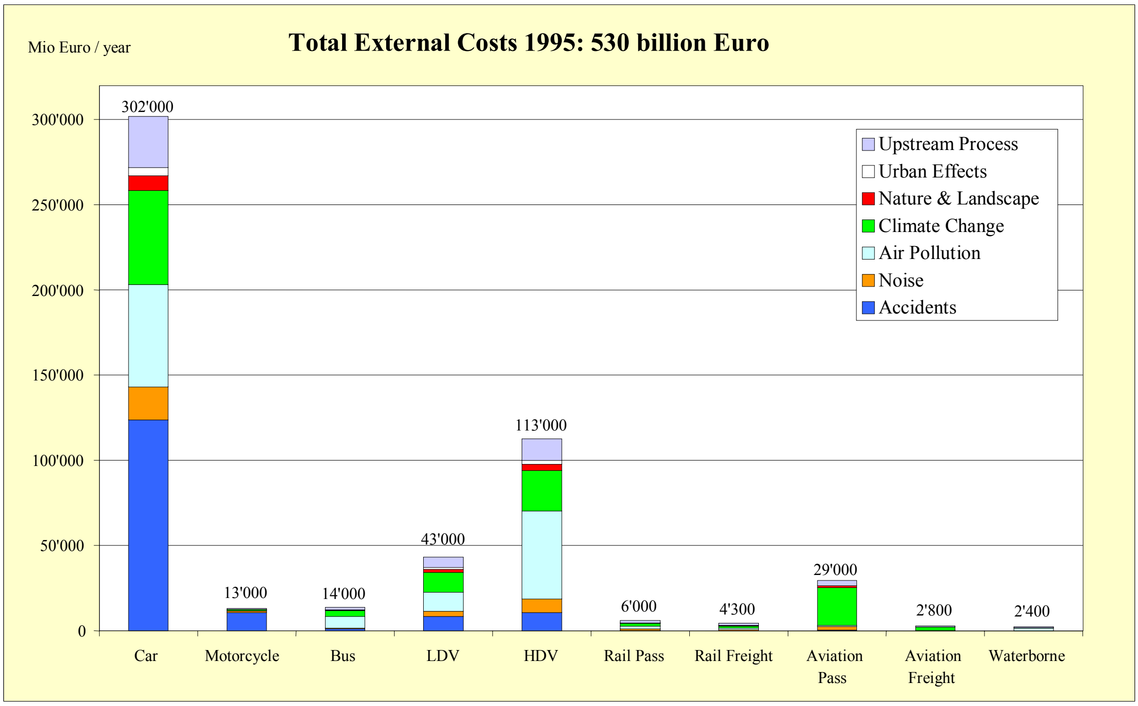
No general estimates have been calculated for Public Service Obligations nor for impact on Economic Development. The literature has concentrated on environment, congestion and accident costs. We begin with the most comprehensive European study, Infras-IWW (2000). Here is a selection of findings, with comments on the methodology as appropriate: costs (per traffic-unit*km) are calculated. Table 12 shows some results of the study, along with methodological comments as appropriate. Figures 4 and 5 give a graphical representation of total costs and average costs per passenger, by transport mode. Figure 6 gives freight figures for comparison purposes.

Table 12: Various externalities as percentage of total transport costs

Type of effect	% Share of total costs (EUR17.95)	Cost components	Most important assumptions
Accidents	29%	- Additional costs of medical care - opportunity costs to society - suffering, grief.	- A value of human life of 1.5 million Euro is considered. - Average costs are equal to marginal costs. There is no specific relation between vehicle-km and accident rates assumed. - Insurance payments are considered in order to estimate external cost components.
Noise	7%	Damages (opportunity costs of land value) and human health.	- The valuation approach is based on a willingness to pay for silent space above 55 dB(A). - Average costs are estimated by a top-down approach based on ECMT data. - Marginal costs are estimated by a modelling approach.
Air pollution	25%	Damages (opportunity costs) of - human health - material - biosphere.	- The results are based on a new and consistent data basis for emissions for all countries (TRENDS/Eurostat). - Health costs are based on a WHO study estimating health costs for France, Austria and Switzerland. - Building damages, crop losses and forest damages are based on results of Swiss expert studies. - Marginal costs are computed by the ExternE model. In order to be compatible with the top-down approach for total and average costs, building damages are adjusted.
Climate change	23%	Damages (opportunity costs) of global warming.	- The data basis is TRENDS. - Unit cost of 135 Euro per tonne of CO_2 assumed. - Marginal costs assumed equal to average variable costs. - The unit costs of air transport are doubled in order to consider the specific risks of emissions in higher altitudes.
Nature and landscape	3%	Additional costs to repair damages, compensation costs.	- A repair cost is used, estimating the desealing costs for different types of infrastructure. - A reference level (unspoilt nature) of 1950 is assumed. - Effects not relevant for social marginal costs, since these costs are infrastructure related.
Separation in urban areas	1%	Time losses of pedestrians.	According to the methodology used in Germany (EWS), time losses are estimated based on random samples of different type of cities.
Space scarcity in urban areas	1%	Space compensation for bicycles.	- According to the methodology used in Germany (EWS), time losses are estimated based on random samples of different types of cities. - The effects are not relevant for social marginal costs, since these costs are infrastructure related.
Additional costs from up- and downstream processes	11%	Additional environmental costs (air pollution, climate change and risks).	- Based on the energy consumption, additional costs for precombustion, production and maintenance of rolling stock and infrastructure is estimated. - For nuclear risks, a shadow price of 0.035 Euro per kWh is assumed, based on willingness-to-pay studies for risk aversion.
Congestion	not taken into account for %	External additional time and operating costs.	- Use of a traffic model to compute marginal and average costs; time values derived from EU research (PETS). Three approaches: - Net welfare loss for road transport facing an optimal congestion tax; - Revenues of an optimal tax; - Time losses relative to a better level of service.

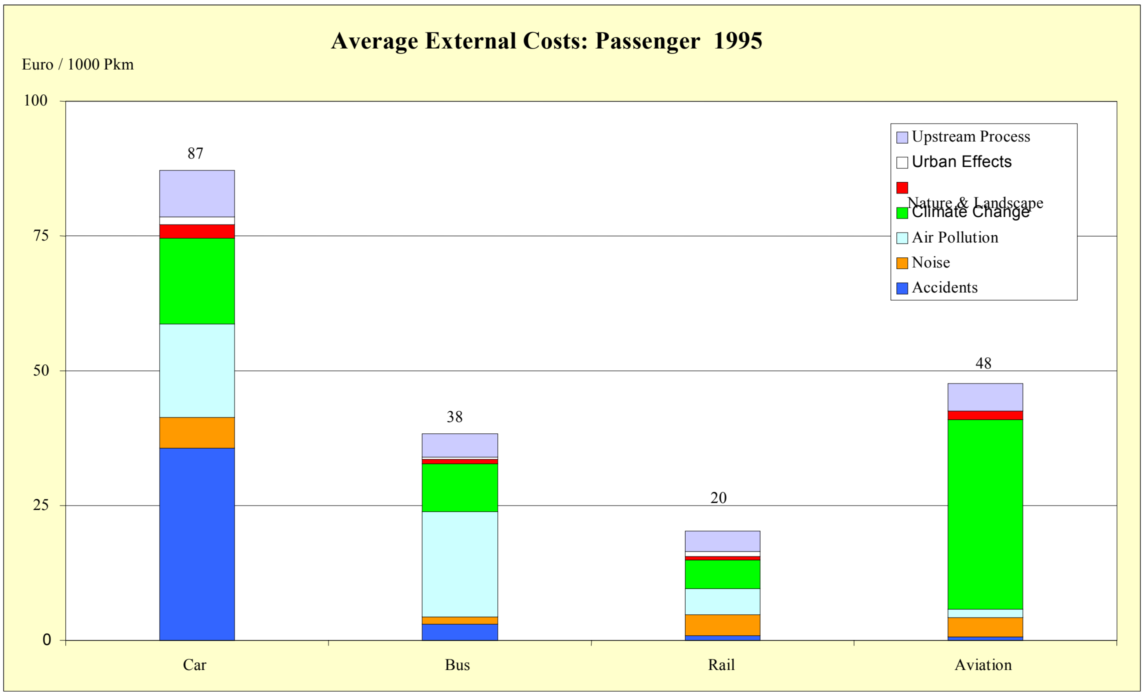
Source:

- PETS 1997: Pricing European Transport Systems-Internalisation of Externalities (deliverable D7); project of the European Commission, Brussels.
- TRENDS 1999: Development of a Database System for the Calculation of Indicators of Environmental Pressure Caused by Transport, Final report of phase 1 and interim report of phase 2, Thessaloniki.
- WHO 1999: Health Costs Due to Road Traffic-Related Air Pollution, an Impact Assessment Project of Austria, France and Switzerland, economic evaluation, technical report, London.
- ExternE/IER *et al.* 1997: External Costs of Transport in ExternE, EU Joule III.



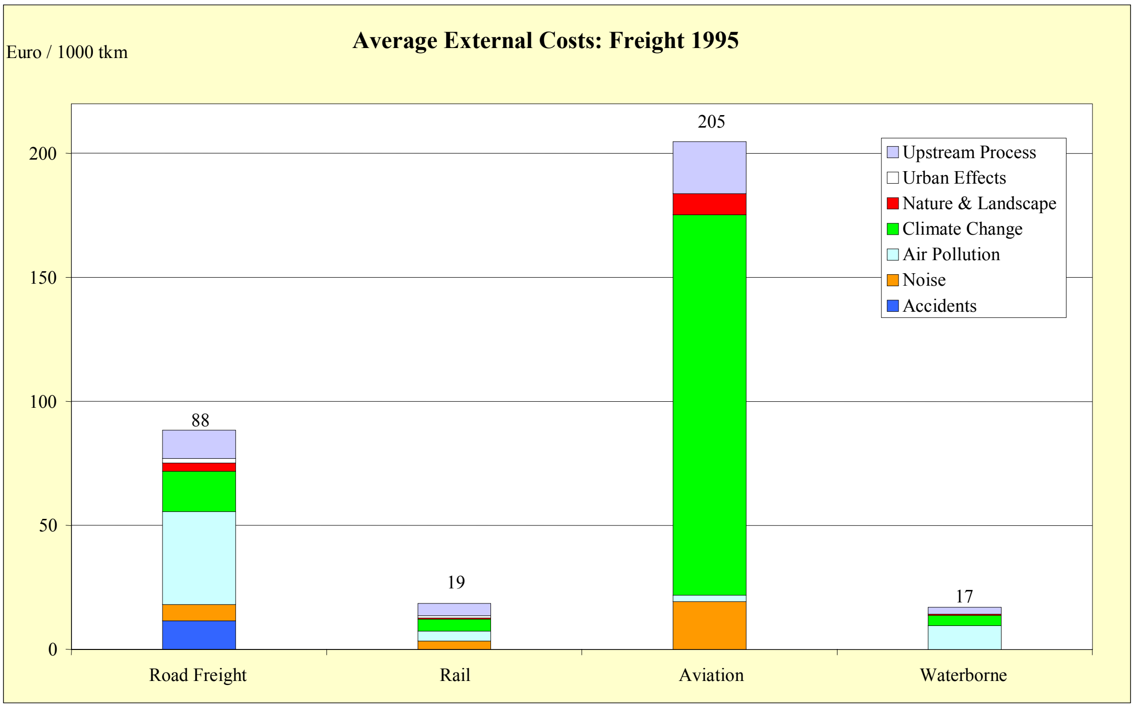
Source: Infras-IWW (2000).

Figure 4: Total external costs of transport 1995 (EUR 17) by transport means and cost category



Source: Infras-IWW (2000).

Figure 5: Average external costs 1995 (EUR 17) by means of transport and cost category for passenger transport (without congestion costs)



Source: Infras-IWW (2000).

Figure 6: Average external costs 1995 (EUR 17) by means of transport and cost category for freight transport (without congestion costs)

Other significant European studies include Amici della terra, Ferrovie dello Stato (2002). One of the most comprehensive American studies is by Litman (1999). For reasons of space we omit a detailed discussion of such findings here, but they are qualitatively consistent with those we have discussed above.

What can we conclude from this massive scientific research effort? It is difficult to summarize the previous numerical results in a unique comprehensive table: the items calculated are not similar, the years are different, the break-down according to modes and the vehicles are not the same, nor are the geographical areas. It is clear, however, that the range of estimates is wide. This is for several reasons:

- There is a high level of uncertainty in some exposure-response functions, such as for global warming, or for air pollution (we do not know well the long term consequences of air pollution, for instance the degree to which it induces cancers).
- There is much uncertainty in monetary values, for instance in the value of human life.
- Many impacts are specific to the situation, especially for noise, congestion, landscape, and ecosystems.

Nevertheless, several broad conclusions can be drawn:

i) Rail externalities are much lower than road externalities, and to a lesser extent, than air transport externalities.
ii) Local air pollution and global warming have roughly the same magnitude, and are a bit higher than noise (except for countries using a high proportion of nuclear power).
iii) Congestion and accident costs, though sensitive to the method of calculation, are generally more important than environmental costs. Environmental costs are significantly lower than private monetary costs.
iv) Total travel time costs are roughly of the same magnitude as private monetary costs.

We turn now to consideration of the conditions of competition in the industry.

3. CONDITIONS OF COMPETITION IN THE INDUSTRY

3.1. Vertical and horizontal integration

3.1.1. Vertical integration

In all network industries, vertical disintegration is a key tool for reforming old utilities. In railways, vertical disintegration is to be viewed as the separation of infrastructure from operational services[6]. As compared with the traditional model of railway organization where a single firm is in charge of both the fixed infrastructure and the rolling stock management, in vertical disintegration competitors are allowed to propose rail services. The infrastructure remains under the control of a public or private monopolist (which requires some public regulation), but market forces are supposed to be strong enough to generate efficiency in services provision. As in other network industries, the dilemma lies in the organization of the interface between the two separated layers. There exists a strong need for coordination between the infrastructure manager and the users of the infrastructure; this is in favor of integration. And conversely, there is a strong need for competition in services which pleads for disintegration. This probably explains why in most countries where competition has been introduced into rail transport, the solution is "partial disintegration" (see Figure 7).

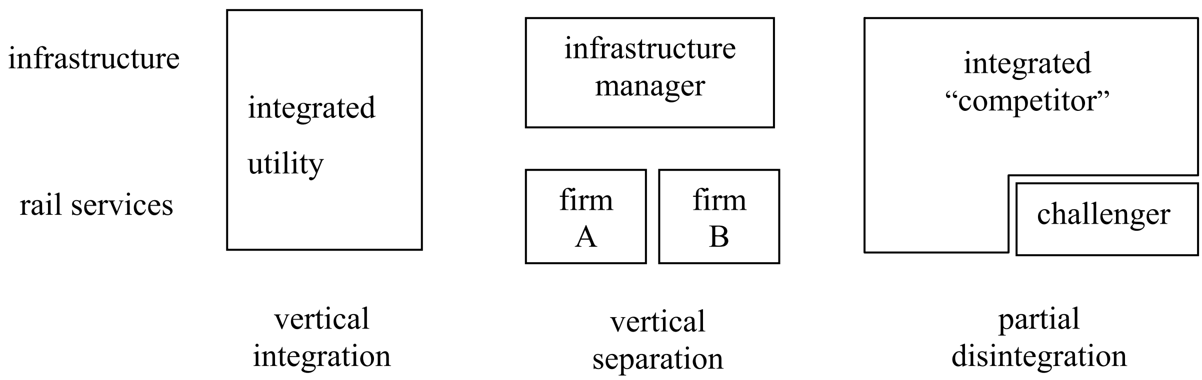


Figure 7: Alternative vertical organizations

When vertical separation is complete, the main problem is to be sure that the monopolist does not abuse its position: it must be regulated. The partial disintegration case is more tricky, since the entity in charge of the infrastructure is simultaneously a provider and a competitor to its challenger. Consequently, it may have some incentives to distort competition in rail services (see Rey *et al.*, 2001) and the public authority faces a complex problem of combined sectoral and competition regulation.

Depending on the nature and the closeness of the integration between upstream and downstream activities, it may also be more difficult for the authorities to have access to the

[6] Note that it is not the only type of vertical disintegration. When a trip form A to C necessitates a stop at the intermediary node B, the segment AB can be viewed by the BC operator as a necessary input to provide AC and similarly, the segment BC is essential for the AB operator to provide AC. For this reason, the separation of AC into two products (namely AB and BC) can be considered as vertical disintegration.

information required for effective regulation than in the disintegrated case – information about costs, for example. However, a well-known advantage of vertical integration is its diminished incentives for double marginalization, so it may be that some kinds of anti-competitive behaviour become less likely under integration even though the authorities' ability to monitor them is diminished.

What does the empirical evidence show about the effects of vertical disintegration on operating costs? Ivaldi and McCullough (2001) test for cost complementarity in freight transport between Infrastructure and operations for USA railroads using a translog specification. According to the estimates shown in Table 13 below, the marginal cost of inter-modal and bulk operation increases with infrastructure output. The negative result for general freight is not statistically significant.

Table 13: Cross elasticity of Marginal Cost

Freight Activity	Infrastructure (t-ratio)
Intermodal	0.31 (1.33)
Bulk	0.52 (2.62)
General	-0.04 (-0.55)

Source: Table 8, Ivaldi and McCullough (2001).

A later article by Ivaldi and McCullough (2002) tests for sub-additivity in the cost function for infrastructure and freight operations. The results indicate that firms running each activity separately have 2.42% higher operational costs than a vertically integrated firm. A study by Cantos (2001) undertakes a similar approach to Ivaldi and McCullough (2001) for European services. Using a translog cost function, the author analyzes economies of scope between infrastructure output[7] and transport operations (passenger and freight) for 12 major European railways along the 1973 –1990 period. The main finding is that the marginal cost of passenger output is increasing with the level of infrastructure value. The opposite result is obtained for freight operations. As Table 14 shows, the cost anti-complementarity in passenger transport holds for all firms, being more severe for the smaller networks.

Other evidence comes from Mizutani and Shoji (2001), who studied the case of Kobe-Kosoku Railway in Japan. They found that vertically separated firms cost 5.6% more than an integrated system[8]. Shires *et al*. (1999a) compared the cost of the Swedish operator after a reform involving vertical separation, and found that operating costs had been reduced by 10%. However, it is difficult to know to what extent such reductions were due to vertical separation per se rather than to other aspects of the reforms.

[7] The monetary value of all infrastructure facilities (track, buildings, stations, etc) is employed as a variable for measuring "Infrastructure Output".

[8] As reported in Mizutani and Nakamura (2001).

Table 14: Cross-Elasticities of Marginal Operating Costs with respect to Infrastructure

Firm		CROSS ELASTICITIES	
		Passenger Marginal Cost Respect to Infrastructure	Freight Marginal Cost Respect to Infrastructure
BR	(UK)	0.119	-0.052
DB	(Germany)	0.076	-0.143
DSB	(Denmark)	0.132	-0.053
FS	(Italia)	0.905	-0.081
NS	(Holland)	0.156	-0.027
NSB	(Norway)	0.133	-0.076
OBB	(Austria)	0.063	-0.116
RENFE	(Spain)	0.082	-0.099
SJ-BV	(Sweden)	0.145	-0.065
SNCB	(Belgium)	0.106	-0.073
SNCF	(France)	0.070	-0.138
VR	(Finland)	0.118	-0.091
	Average	**0.108**	**-0.085**

Note: All values are statistically significant at 5%
Source: Table 5 in Cantos (2001).

In general the evidence, such as it is, is mixed and inconclusive. All studies except one (the Shires *et al*. study of Sweden) estimate cost complementarities using data from currently integrated firms, which leaves the studies vulnerable to bias due to internal cost-allocation rules, and which means they are unable to take account of what may be the most important effects of vertical disintegration, namely transactions and coordination costs. The UK case study in Appendix 1 below suggests these costs may be large, though it is difficult to generalize from a single (and rather unusual) case. Overall, it seems safe to conclude that existing cost studies do not show that vertical disintegration of infrastructure from operations imposes an important cost penalty. However, given that such separation has occurred very rarely to date, the value of such studies in predicting the future consequences of such separations would appear to be very limited.

One important issue on which such studies cannot realistically shed light concerns the role of vertical coordination in influencing the evolution of network structures. In airline networks (unlike in rail networks), market entry can create new routes without the need for prior infrastructure investment. To be more precise, provided airport infrastructure exists at the cities at either end of a route, any entrant to the industry can create a direct flight link between two cities where none existed before. However, this cannot happen in railways, where tracks need to be laid before trains can pass. Such entry by airlines has proved of immense importance in shaping the evolution of structures towards hub-and-spoke models in the US, and has begun to be important in allowing new entrants to offer competitive services in the European market. Furthermore, although airport infrastructure can become congested and thereby impose a constraint on network development, the creation of new routes is an important mechanism whereby signals of the need for airport infrastructure investment are perceived. In railways, though, network investment will always need to lead rather than lag new route entry by service operators. That implies that the infrastructure operators will need to have much closer coordination (concerning future operation intentions) with service operators than is necessary in the air transport industry. Vertical integration and vertical disintegration with close investment and operational coordination are both feasible options; vertical disintegration with an arms-length relationship between infrastructure and service operators is not.

The literature on "transactions costs" (see Williamson, 1985) has provided some important insights on the role of vertical integration in industries with high sunk costs of investment (as in railways). For obvious reasons, it is important to ensure that productive investment does not fail to take place because of a lack of coordination of the upstream and downstream parties' intentions, due to their lack of integration. Williamson's insight is that such failure may not occur simply because of a breakdown of communications but for a much more fundamental reason, which he terms the "hold-up problem". Suppose that one party invests prior to the other, and that the investment creates a "specific asset" – one that is worth much less outside the relationship between the two parties. For instance, the asset may be a stretch of railway track that is adapted for high speed trains, which only one operator can run (other operators can run normal trains which do not make full use of the valuable track). Then, as soon as the investment has been irrevocably committed, the HST operator has an incentive to toughen its bargaining position, threatening not to make its own share of the relevant investment. The track operator would have, in effect, to bribe it to invest by lowering the access price to the track, and it might have to lower the price all the way to the price it could charge other, non HST operators. Naturally, the fear that this might be the outcome would be a disincentive to investing in the track in the first place.

Various possible solutions to the hold-up problem have been proposed, including long-term contracts (which in this example would set the access price at an agreed level even before the track investment had been committed). Long-term contracts can be difficult to write, however, especially when future circumstances may change in unforeseen ways. Instead, vertical integration between track and service operators may resolve the problem by ensuring that neither has the incentive to bargain with the other after the commitment of the investment. Integration does not have to be complete; joint ventures on specific projects by partners that otherwise remain separate are an alternative that may work when the projects are sufficiently distinct. Nevertheless, it is worth bearing in mind that ill-considered vertical disintegration by regulatory fiat may cause difficulties of investment coordination that are not just "communications problems" but go to the heart of negotiation incentives.

To summarize, vertical integration has some disadvantages in a transport network, due to the potentially greater opacity of costs and other operating information that makes effective regulation more difficult, and leads to a risk of anti-competitive discrimination by the network operator against services supplied by a downstream competitor. However, the list of potential advantages of vertical integration is long. It includes some aspects on which empirical evidence is available (notably the extent of vertical economies of scope), and others (notably transactions costs and the risk of hold-up problems) on which evidence is scarce but which may plausibly be extremely important. The overall balance of advantages in vertical network integration is therefore a subject on which further information and research is very much required.

3.1.2. Horizontal integration

Passenger rail transport involves a number of distinct activities, some of which may be more open to horizontal disintegration than others. We distinguish them as follows:

Provision of services on existing infrastructure linking a given city-pair

Competition in this activity is sometimes known as "competition on the tracks", and it is rare, though one or two instances of it have been seen in Germany in recent years. It is interesting to note that the Prussian rail system in the first half of the nineteenth century allowed for the possibility of competition on the tracks, although this possibility was not taken up (see Millward, 2002). After the nationalization of the industry in 1879 such possibilities were abandoned.

Provision of services on existing infrastructure in adjacent regional markets

As footnote 6 noted, it is far from clear that such a structure deserves to be called vertical disintegration at all, since services between adjacent regional markets are complements, not substitutes for one another. This implies that it is unlikely that the presence of rail services in adjacent markets will exert much competitive pressure on prices or service quality, except in so far as the terms of renewal of franchises takes comparative performance into account. This is the model that has been adopted in the UK (see Appendix 1) through the use of franchising contracts.

Leasing of rolling-stock

In principle there seems no reason by there should not be effective competition in the ownership and leasing of rolling-stock. The UK rail reforms of the 1990 were designed to bring about such competition by creating three ROSCOs (Rolling Stock Operating Companies). However, there has not been effective entry into this market, and it is likely that barriers to entry are significant, due to the long lead times for investment and the difficulties in transferring rolling-stock between countries (which gives investment costs a significantly sunk character).

Timetabling and reservation services

In the past such services have been considered as something of a natural monopoly, though with the development of the internet this may be changing. At all events, low-cost entrants in the airline industry have thrived in spite of not sharing the distribution and booking systems (such as Amadeus) belonging to the major flag carriers.

Infrastructure provision on a given city-pair

Historically there were a number of instances in the 19th century where rival companies would build routes linking two important cities, and such competition has continued to take place in Japan to this day. In all such cases the infrastructure provider was also the provider of services on the infrastructure concerned, and no cases are known where there was competing infrastructure on a city-pair without vertical integration.

Infrastructure and service provision in adjacent regional markets

For the reasons given above, this is best characterized as a system of parallel monopoly rather than horizontal competition. Where it appears to have worked well it has either been accompanied by strict yardstick competition (as in Japan, see Appendix 1), or by strong inter-modal competition (as with trucking for the US freight transport networks).

The main message seems to be that horizontal competition in the strict sense is extremely rare in railways, and it is very hard therefore to judge what would be the likely consequences of mandating it on a large scale. Parallel monopoly, which is what has been tried in different ways in the UK, the US and Japan, is not at all the same thing.

3.2. Service diversification and economies of scope

Diversification is important for the analysis of competition for two reasons. First, it may affect costs if there are significant economies of scope between the activities. Secondly, diversified firms may be able to cross-subsidize their markets. This may distort competition since single product firms cannot divert resources from other markets[9]. Diversification is hard to quantify. There exist some well-known cases of train operators who are also sellers of air transport, disks and colas. There is also the possibility of using the infrastructure for other activities, such as high-voltage electricity cables or broadband communications. But it is risky to diversify too much: the positive portfolio effect can be overcompensated by negative fuzziness effects (loss of control, diseconomies of scope). Similarly, attempting to differentiate services very precisely by customer may risk increasing costs because of the loss of opportunities for pooling; for instance the distinction between first and business class travel has not proved economic to implement in Europe, and the number of classes of rail travel has declined since the nineteenth century.

We are not aware of any studies directly estimating such economies of scope in railways. However, one implication of the findings on economies of density reported in section 2.3 above is that operators offering only point-to-point services are likely to be at a cost disadvantage compared to rivals that can offer at least a small network of routes as "feeders" (producing what is called "intraline" traffic). To put the point another way, point-to-point entrants may need to have a substantial advantage in direct operating costs to be able operate profitably against a network operator. The extent to which such competition is possible will depend on the characteristics of the routes in question: for instance, low-cost airlines in Europe such as EasyJet and Ryanair have overwhelmingly concentrated their entry strategies on routes where there is substantial point-to-point traffic and they are unlikely to be disadvantaged by their inability to intra-line. It seems likely that any low-cost competition on the tracks in European rail services in future will similarly concentrate on routes with high point-to-point traffic, which tend to be profitable for incumbent operators not so much because of high prices as because of their ability to benefit from economies of density.

In order to prevent cross subsidization, governments promulgate rules that oblige some regulated firms to publish separate accounts. This obligation has the advantage of transparency. But it does not prevent multi-product firms from allocating common costs to activities where competition is weak. The purpose of such cost allocations can be to commit the firm to a very aggressive price policy (though one still above cost) in markets where competition is tough. If such a policy is effective it may not even need to implement aggressive pricing provided competitors get the message and themselves use an accommodating strategy, or even stay out of the market altogether.

Almost certainly the presence of potential cross-subsidy will be as important as the presence of point-to-point demand in influencing the evolution of competition on the tracks if and when it becomes more common in Europe. Carriers that can reduce their costs on one route by sharing rolling stock or marketing services with other routes (including routes in other countries) may be at a considerable advantage compared to carriers that cannot, whatever regulators may do to try and level the playing field. Similarly, firms that can obtain subsidies for the use of rolling stock on some routes may be able to benefit from such subsidy to compete more aggressively on other routes.

3.3. Inter-modal competition

For the transport of passengers, train companies compete mainly with buses and cars (taxis or private cars) on short and medium distances, and with airline companies for long distance, in particular for interstate trips.

For the economic analysis of this inter-modal competition, it is useful to observe that:

- on the supply side, it is competition between heterogeneous technologies;

[9] Rey *et al*. (2001) emphasize that even if such cross-subsidization is possible it may not be in the interest of the firm to do so.

- on the demand side, the competing products are more or less differentiated;
- the competing modes are regulated following different rules.

When firms compete with very different technologies, it is rare from them to provide exactly the same product. The electricity industry is an exception - the consumer of 1kW is totally unable to say if it has been produced by a gas plant, or a nuclear plant, or a hydro plant.

We provisionally suppose that the product is simply defined in terms of a trip between two given nodes, starting and arriving at given times. There are obviously numerous different possibilities to be considered, but here we limit the analysis to the case of starting and arriving points in the center of two towns since this is the interesting case for the train companies.

What are the advantages of a bus company for the same service?

- more flexibility in terms of time departure and arrival;
- more flexibility in terms of path;
- lower fixed costs;
- the optional possibility to "collect" and to "distribute" passengers around the nodes instead of at the nodes.

What are the comparative drawbacks of the bus solution?

- lower comfort;
- possibly lower safety (depending on the kind of route);
- air pollution and road congestion.

If passengers are mainly interested in being transported from one town to the other, the products supplied by the train company and the bus company are merely differentiated vertically by the duration of the travel: the shorter, the better. In this case, the train can be seen as a high-quality good and the bus as a low-quality good. From this point of view, on medium distances (Paris-Lyon, Paris Marseille), the train has long been viewed as of low quality with respect to plane but thanks to high-speed trains, the "quality gap" has vanished. It has even switched to the advantage of train companies for the shortest segments (Paris-Lyon).

Vertical differentiation can also be defined in terms of safety, reliability and comfort. Unlike buses, trains and aircraft can propose different classes in order to discriminate between passengers on the basis of their willingness to pay.

The main difference between trains and cars or buses is in terms of horizontal differentiation[10]. Because the demand is extremely differentiated in geographical terms, road transport can propose products that are "closer" to the individual needs. Indeed, unlike road transport, trains need complementary modes between the individual departure and arrival nodes on one hand and the stations on the other hand (see Figure 8). The same applies to airlines.

[10] The distinction between horizontal and vertical differentiation is that vertical differentiation is with respect to properties of the service that all consumers can be presumed to value (such as speed), even if they value them to different degrees. Horizontal differentiation, however, is with respect to properties of the service which are ordered differently in the preferences of consumers, such as physical location (where one consumer prefers location A to location B, another the reverse).

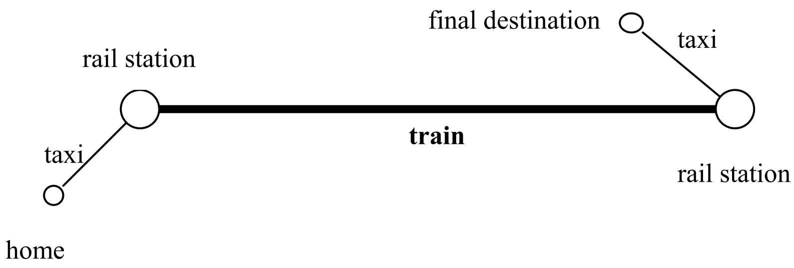


Figure 8: Multi-modal transport links

The regulation of competing modes of transport is not subject to symmetric rules. Different technical and economic agencies are involved in controlling the operation and the development of the different types of transport.[11] For example, when fatal accidents occur in rail transport, the demand for increased security is much stronger than if the same number of people is killed in car crashes. Similarly, pollution by trucks and the damage they create on roads may not be calculated in a way that is comparable between rail transport and road transport.

The intensity of inter-modal competition can be measured using cross-elasticities of demand when these are available. For passenger rail travel as a whole, the findings reported in section 2.2 implied that these were low in the short term. However, in the longer term, as well as for particular markets such as large cities less than 250 km apart, competition may be stronger. For instance, inter-modal competition between rail and air is quite strong for such cities, as the experience of the Paris-Lyon, Paris-Marseille and Madrid-Seville high-speed train services indicates.

However, it is useful to remember that rail travel can be complementary to other modes (indeed it typically depends on other modes to allow rail travelers to complete their journeys). To put matters another way, for train operators, the other modes of transport can be their best allies as well as their fiercest competitors.

To illustrate this question, we present now a model of regulated inter-modal competition.

The model of Brauetigam (1979) can be readily adapted to passenger transport. We consider the market for transporting passengers between two towns. There are m modes competing in this market. Trains are mode 1. They are characterized by economies of scale with the consequence that marginal pricing would create financial losses. The other modes (private cars, taxis, buses …) exhibit decreasing returns to scale and there is free entry in these modes. Consequently, one natural monopoly (mode 1) is competing with $m-1$ competitive modes.

There are $j = 1,\dots,n$ different types of passengers. The number of passengers of type j transported by mode i is denoted x_i^j. Intra-modal services are perfectly homogeneous for one class of passengers (for example, all taxis are the same). The aggregate utility of type-*j* passengers is:

$$u^j(x_1^j,\dots,x_m^j) \qquad j = 1,\dots,n$$

[11] Brauetigam (1979) examines "how second best prices might be set if one of the modes has scale economies (and railroads appear to be the most likely candidate) and the other modes (water and motor) do not."

This formulation presumes that each additional type-*j* passenger on a given mode has a positive utility, that this positive utility is larger than the disutility created to former type-*j* passengers, and also that there is no positive or negative externality between different groups of passengers. Some "discomfort" parameter could be added to the model.

Passengers have no market power. One can derive the set of marginal conditions for the demand of transport by type-*j* passengers:

$$\frac{\partial u^j(x_1^j,\ldots,x_n^j)}{\partial x_i^j} = p_i^j \qquad i = 1,\ldots,m$$

There is intermodal service differentiation since trains and competing modes differ in terms of speed, comfort, reliability etc. This assumption can be modeled by

$$\frac{\partial p_i^j}{\partial x_i^j} \le \frac{\partial p_i^j}{\partial x_{i'}^j} < 0 \qquad i \ne i'$$

In other words, for type-*j* passengers, modes i and i' are weak gross substitutes.

Starting from these assumptions, one can determine:

- Second Best allocation, i.e., how inter-modal competition distorts the optimal activity of the train company that has to balance its budget. It appears that even the "competitive" (but regulated) modes need to depart from marginal cost pricing.
- Partially Regulated Second Best, i.e., what is the second-best number of passengers by train when the social planner does not control the use of the other modes of transport.

This model is rather old-fashioned, but it gives a good introduction to the problem of "unfair" competition between heterogeneous technologies with the intervention of public decision-makers that interfere with market mechanisms.

3.4. Access to infrastructure and barriers to entry

Infrastructure refers to what are sometimes called "essential facilities" - those aspects of a network that would be seriously costly for other rail service operators to duplicate. In practice in rail transport this includes track, signalling, stations and possibly some maintenance facilities (note that the maintenance labour force is not, properly speaking, a part of infrastructure). It may also consist of some ticketing and distribution facilities, though the costs of duplicating these have been drastically reduced in recent years by the internet.

As long as the industry remains completely vertically integrated, barriers to entry are "natural", in the sense that entry is limited by the huge fixed costs necessary to install new tracks, stations and all the software to operate them. The greater the vertical disintegration, the easier is entry likely to be. This consideration even prompted the UK authorities to negotiate artificial barriers to entry with franchisees in order to ensure that on-track competition did not lead to increased demands for subsidies (see Box 4). However, such entry is likely to be limited to routes where existing infrastructure is in place.

Lowering economic barriers to entry should not mean decreasing quality and safety. For this reason, governments define severe institutional barriers. Entry is subordinated to obtaining a license or to signing a contract that describes in detail the technical, economic and financial requirements of the deal.

With this type of barrier in place, the risk is that governments inefficiently protect domestic firms, mainly the incumbent, advocating safety reasons as a means of denying entry to foreign firms.

Finally, barriers to entry can be strategic, that is voluntarily erected by one or several incumbents in order to blockade access to the market by newcomers. Such strategic barriers may take many forms, from marketing expenditure to investment that lowers marginal costs and commits incumbent firms to aggressive pricing behaviour.

Box 3: Barriers to entry in the UK railways

The original privatisation plan envisaged two types of access agreements for passenger services: franchised passenger services access agreements; and open access agreements where operators would be able to negotiate open access agreements with Railtrack. It soon became clear that encouraging on-track competition would conflict with the government's objective to reduce subsidies.

Thus the Rail Regulator has devised a regime of 'moderation of competition' for the passenger franchises. This regime determines in advance the degree of exposure to competition which each franchisee could face, ensuring that the adverse impact on the franchising process would be minimised. New entry is restricted through contractual control over Railtrack's ability to sell access rights to any operator in addition to those contained in their initial access agreements. The mechanism was intended to allow the market to determine as far as possible where competitive entry should occur. The protection given was to be over the markets which operators serve rather than the track on which they operate.

The mechanism operates in two stages. The first started when the last British Rail TOC has its long-term track access agreement approved and was originally planned to expire on 31st March 1999. The second was planned to expire on 31st March 2002.

First stage of open access:

In the first stage, franchise holders can nominate a list of substantial point-to-point flows on which no new entry for scheduled passenger services will be permitted without the franchisee's agreement. Franchise operators can nominate additional flows, not conforming to this rule, subject to the approval of the regulator. This is designed to allow passenger operators additional rights where there is no conflict with the protection given to other operators.

Second stage of open access:

In the second stage, some new competition via open access is intended to be introduced. As with stage 1, the key building block is the station to station flow and franchisees will be expected to re-nominate flows, but in stage 2, franchisees will only be allowed a partial protection from new entrants. Up to a threshold level set by the Regulator, new entry can then occur, and compete for the operator's business. In stage 2, the competition mechanism is designed to limit the amount of the operator's passenger revenue that is open to competition from another operator.

Since the most important barriers to entry arise from the long-term nature of investment, both in infrastructure and in rolling stock, we consider this issue separately in section 3.5.

3.5. Infrastructure investment and financing

As in other parts of rail activity, a distinction has to be made in investment issues, between infrastructures and rolling stock. This distinction is important for several reasons:

- The life of investments is very different: from ten to thirty years for rolling stocks, several centuries for infrastructure if it is properly maintained.
- The concern of public authorities for infrastructure: for instance strategic and geopolitical concerns, as well as the need of public intervention to obtain the right of ways to achieve a new infrastructure (however, it should be noted that much rail infrastructure was privately built as well as operated during the 19th century in Europe).
- The fact that infrastructure has many characteristics of indivisibility: a new rail link needs large amounts of money in order to be useful, while a new car alone may be useful.
- The sunk cost characteristics of infrastructure: when the investment has been made, it cannot be given another use, while rolling stock can be resold (through the second hand market).
- The fact that the Commission of the European Union has launched a policy towards a separation of infrastructure and rolling stock operation. In all European countries there is at least an accounting separation, and in most countries an organisational one. This organisational feature is specific to Europe; in no other part of the world has railway reform endorsed such a separation (in most countries, the separation has been a horizontal, not a vertical one). After this separation, it appears that infrastructures remain in the hands of the public authorities, while rolling stocks are operated by firms that are more and more privately managed.

Let us examine successively the economics of rolling stock and of infrastructure.

3.5.1. Rolling Stock

The economics of rolling stock is largely independent from the public authorities and belongs to the market economy. But this market is different in the various parts of the world and it is interesting to relate the features of the North American markets and the features of the European market.

In North America, there are a large number of rail operators, both for freight (long distance transport) and for passengers (short distance transport). There are also a rather large number of rolling stock providers, apart from the two big firms, General Motors and General Electric. So the market is fairly close to a competitive market, with a large number of suppliers and of customers and a rather vivid competition, with a tendency to outsourcing. For instance the two big makers subcontract about 80% of their turnover.

The picture is quite different in Europe, or more precisely was quite different, because things are changing and the situation is moving closer to the American one. Around 10 years ago, the railways operators of the main European countries had no link between them, were state monopolies, and had no obligation to use international competition for the provision of rolling stock. In each country, there was a bilateral monopoly between the rail operator (DB in Germany, SNCF in France, FS in Italy) and the rolling stock provider (Siemens in Germany, Alsthom in France, Ansaldo in Italy). The connection between the rail operator and the rolling stock provider was very tight in each country, without any link between one country and another. The technical staff cooperated closely, and the results were:

- Fast technical progress, which led to the various HST of each country (ICE in Germany, TGV in France, Pendolino in Italy – the last being strictly a tilting rather than a high speed train), which needed close cooperation between the future operator and the maker.
- Small production runs, leading to high costs.
- Standards different from one country to another. Although the width of rails is the same, except for Spain and Portugal, the loading gauges, the power and the signalling systems (including the official language in use between the drivers and the controllers, which is the domestic language) are different from one country to another, and there are even significant intra-country differences.

In contrast, the outcome of the American market structure was quite the reverse: rather low technical progress, long production runs, widely used standards, interoperability of rolling stock over the whole continent.

Things are changing in Europe: under the pressure of the liberalisation and the Commission efforts to widen competition, tendering procedures are becoming international, the opening of the market to international competition has induced American providers (such as Bombardier) to enter the market through direct sales to European rail operators or through mergers and absorption of European firms. Foreign makers are becoming active in each domestic market. EWS, the UK freight train operator, buys rolling stock from a Chicago firm, Thrall. Many European rail operators pay more attention to the buying strategy and to the prices they pay. They induce competition, and do not hesitate to buy foreign materials such as SNCF, or DB. There are also vehicle leasing companies, notably in the UK, though this is also beginning in Germany.

An interesting features is the kind of competition which is evolving in the HST market: we see that in the overseas markets, each of the three main providers (Siemens, Alstom, Fiat) are acting as competitors (case of Korea), while on the European markets they act cooperatively, as can be seen from the international links such as Paris-Bruxelles-Cologne or the Eurostar.

There has been little systematic study of the economics of rolling stock manufacture or investment, nor much precise and detailed statistical information. In particular, there is no official comprehensive record of rail rolling stock investment and no general figures can be given concerning its evolution. Nevertheless, Affuso and Newbery (2000) note that the immediate aftermath of privatization in the UK was a collapse in rolling stock investment (though it had been on a downward trend beforehand), though their econometric study does not find grounds for attributing this to insufficient length of franchising contracts.

3.5.2. Infrastructure

In contrast to rolling stock, the economics of infrastructure is not much driven by market forces, but much more by the policies of public authorities. This is probably the reason why so many econometric studies have been made on this subject.

These studies are related to:

- infrastructure costs and the level of economies of scale and scope;
- the decision process for the implementation of new infrastructures;
- financing issues and especially the role of private financing.

Infrastructure costs in a broad sense can be listed as follows:

- building costs;

- maintenance and operation costs;
- external costs, and especially environmental costs, congestion costs, and unsafety costs.

Building costs are rather well known through various project studies in many countries. Table 15 provide some average values for France, comparing rail costs and those for other modes. These costs are averages, and the real cost depends on specificities such as the climate, the topography, the density of inhabitants (the more dense is the population, the more expensive are rights of ways and constructions such as tunnels and bridges).

Maintenance costs have been the subject of many studies estimating returns to scale. It appears that maintenance and operations of infrastructure exhibit highly increasing returns to scale, or, put in another way, that the fixed costs of maintenance and operations are high. A simple measure of the degree of returns to scale is the ratio of average costs to marginal costs. When this ratio is higher than one, there are increasing returns to scale. In industries such as telecoms and air transport these ratios have been estimated to lie between about 1.5 and 2.0. The ratio in rail infrastructure activity is around 3 to 4 according to study and country.

Overall, it appears that these marginal costs are a long way below average infrastructure costs: this is true even if externalities are included among the marginal costs. This implies clearly that marginal cost pricing cannot cover infrastructure fixed costs, so the need for public subsidies remains even in when externalities are taken into account.

Table 15: Infrastructure costs for various links, millions of Euros 1995

Mode of transport	Nature of link	Mean cost
Road	1 km expressway 2X2	3.1
	1 km route nationale 3-lane	1.4
	1 km route dptale 2-lane	0.5
	1 km route communale	0.26
	1 km autoroute	6.5
Rail	1 km double electrified	4.3
	1 km single electrified	2.3
	Electrification of 1 km single	1.5
	Quadruplication of double	2.7
	1 km high-speed track	6.5

Source: Quinet (1994).

One of the features emerging from the table is that rail links are somewhat more expensive per kilometre of track than road links. Allied to the fact that road links are capable in principle of supporting a more or less continuous flow of traffic, while rail links may remain empty for significant periods between the passage of trains, this underlines the need for rail links to generate significant traffic levels before they can attain a reasonable social rate of return in competition with road. It should also be noted that all these costs are increasing in real terms over time due to environmental constraints.

3.5.3. Decision-making for infrastructure investment

Public authorities are highly involved in decisions about infrastructure. In all European countries, the infrastructure manager (IM) is either the state or a public firm, and public authorities subsidize the investments.

In Sweden, the IM (Banverket, initials BV) is a part of the administration; investment decisions are taken in accordance with and under the control of the Parliament, which decides the budgetary funds to be granted to BV and has greatly increased the funds since the beginning of the reform process.

In France, the IM is a public firm, with the same status as SNCF, the rail operator. Its funds come from infrastructure charges, which are far below the total expenses, and from borrowing and public subsidies. As in other countries, the debt of the IM is very large.

In Switzerland, the IM is a branch of the CFF and benefits from funds coming from the special tax on road haulage.

In Germany, DB Netz is a branch of the holding DB AG and is still closely linked to the rail operators.

The case of the UK is interesting in that, at the beginning of the reform, Railtrack was established as a private firm under the control of a regulator. However, the system proved to have many drawbacks coming from the difficulties of regulating the system. Recently the status of Railtrack was changed to an organisation equivalent to a public firm.

The involvement of public authorities in infrastructure influences the technical aspects of the decision process. Choices are generally screened by cost-benefit analysis. The main difference between cost-benefit analysis (CBA) and current procedures in private management is that CBA aims at reckoning the whole social profitability and not only the profitability of the IM. This implies inclusion of certain external effects:

- the effects on environment;
- the effects on user's surplus, which are mainly the gains in travel time that cannot be captured by the IM through a sufficiently differentiated charging system;
- the effects on the profits of the other infrastructure managers.

These calculations are performed in Germany, Sweden, France, but not in Switzerland. In the UK, the procedure up to now did not include these external effects, as the investments were decided more or less by the private firm Railtrack, but things will probably change since the nationalization of the IM.

Where they are attempted, these calculations prove to be difficult, uncertain, and easily manipulable: information about future traffic is quite asymmetric between the IM and the regulator, the calculation of users' surplus is uncertain, as is the calculation of the external effects, due to the specificities of each investment and to the difficulty for the regulator in controlling the accuracy of both the data and the calculation process. The result is that there is often a substantial difference between the recommendations of the CBA and the subsequent decisions.

Fragmentation has raised a new problem, by splitting the infrastructure builder and the infrastructure user: a problem of commitment arises, the IM being cautious to build a new infrastructure the cost of which is totally sunk, and becoming a hostage of the rail operators. Long term contracts between the rail operators and the IM are one possible solution, though one that may conflict with the competition process that the European Commission is trying to implement.

The involvement of the public authorities is not due solely or even mainly to the need to take externalities into account, but also by the poor private profitability of most infrastructure investments. In France for instance, the HST from Paris to the East of France and onwards to Germany has a rate of return of only about 2%, and the rate of return of the Alpine tunnels in Switzerland has roughly a similar value. The Betuwe line, which will be built in Netherlands to link Rotterdam and the Rhine mouth to Germany, was originally intended to be privately funded, but this now seems unlikely.

The support from public authorities for rail investments is increasingly driven by the general policy orientations of the Commission and of many countries: they intend to develop rail traffic in order to cope with road traffic growth and negative congestion and environmental externalities. The result, coupled with the macro-economic constraints of short run oscillations, is a substantial degree of fluctuation in infrastructure investments from one year to another.

3.5.4. Use of and Access to Infrastructure

The use of infrastructure is determined first by infrastructure charges and secondly by slot allocation procedures. Infrastructure charges are very different from one country to another in Europe. Their structure is rather similar. Each country has issued regulations about infrastructure pricing. These tariffs (NERA 1998) generally take 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 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 term related to the size of the train.

These principles have taken various forms in each country, and things are changing rapidly. In the UK for instance, infrastructure pricing is presently composed of a lump-sum negotiated term with each RO for the base service and a marginal term for each additional train. In Germany, up to recently there was a menu of two tariffs: a proportional one, fit for small ROs, and a two part tariff, fit for big ones. Recently a change has suppressed this two-part tariff, which was judged by the new entrants as discriminatory.

The infrastructure charges cover different proportions of infrastructure costs in different countries, these proportions being very low in Sweden and France, and significantly higher in Germany and in the UK.

Infrastructure charges do not completely solve short-term adjustment problems. Contrarily to the case of road, there are still applications for incompatible paths, coming from different operators. These problems require path allocation procedures. Little attention has been paid to path allocation issues, probably because it is not a new problem, having already occurred as a problem for the historical incumbents. The European Union has recently mentioned it in some directives, and the recent Directive (European Commission, 2001), which aims at dealing with this question, just fixes the procedures to be applied: no discrimination between the applicants; a precise time-table for the applications; and the allocation should be made by an independent body. The algorithm for solving conflicts is not detailed; in fact all countries have established priority rules which are in the line of the procedures used by the historical incumbents.

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. There are of course country peculiarities. In the UK and Denmark, some subsidies can be granted to combined trains. In Germany and Switzerland, the regulation implies that, within equal priority rank for conflicting applications, the higher bidder is chosen. 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. Besides these dispositions related to the planning procedures of services, real time operations are generally achieved by the IMs through hierarchical relations. In the UK, a performance regime is set up, implying bonuses and penalties for delays and on-time services to be paid between the ROs and the IM.

3.6. Are all facilities natural monopolies?

Network industries are commonly viewed as a combination of activities with different characteristics: for example, in railways we have the management of infrastructure on one side, and the operation of rolling stock on the other. Liberalization consists in opening to competitors the layers of the industry where costs do not exhibit strong economies of scale. Other layers may be better managed as a monopoly (typically to avoid duplication of fixed costs). And competition can be fair only if access to the essential facility is guaranteed to all competitors.

But are all fixed equipment and immaterial assets essential facilities? The answer depends on:

a) the cost characteristics of the isolated piece of equipment;
b) the cost characteristics of equipment in the industry as a whole;
c) the benefits or damages to competition of access opening.

a) If the equipment under scrutiny currently displays decreasing average costs of operation, and if this remains true after the entry of new users, opening access is potentially beneficial for everybody: for the entrant who is not obliged to invest in a new equipment, for the incumbent who benefits from additional decrease in average cost and for society since duplication would be inefficient (Figure 8).

When there are decreasing returns to scale, the access of competitors will increase the average cost of the incumbent (as in Figure 9). The entrants benefit from the investment of the incumbent since the average cost they incur is less than if they had to install their own equipment (AC_e). In terms of the welfare of society as a whole, it depends on a comparison of the total costs incurred by the firms when equipment is duplicated and when it is shared.

Decisions about sharing of facilities will therefore be quite sensitive to details of the costs involved. For example, some repair and maintenance facilities may have the characteristics of a local natural monopoly and should be made available to competitors. Others may be replicable by competitors with little disadvantage in terms of costs (this is more likely to be true when there are no large indivisibilities involved in the construction of such facilities, and when the likely traffic on the competitors' services is large enough to ensure the facilities are reasonably fully utilized).

Similarly, technology may influence the nature of such comparisons over time. For instance, recent developments in information technology make it much easier than it used to be for competitors to replicate the ticketing and distribution facilities of incumbents with little or no cost disadvantage – either through internet booking (as low-cost airlines have discovered) or through small portable facilities on board trains. This implies, that in these activities sharing of facilities needs not to be enforced.

b) Access is likely to be optimal if the facilities in question can be clearly isolated from the rest of the infrastructure. However, some inputs do not perform in the same way when they are operated in isolation as when they are connected to others. To open access to one facility without taking into account the effects on linked facilities can be damaging for the incumbent and for passengers. For instance, granting open access to a repair station because there remains some idle capacity (see argument a) above) can create congestion on surrounding lines and stations with the effect of increase the total costs of the incumbent and the welfare of passengers.

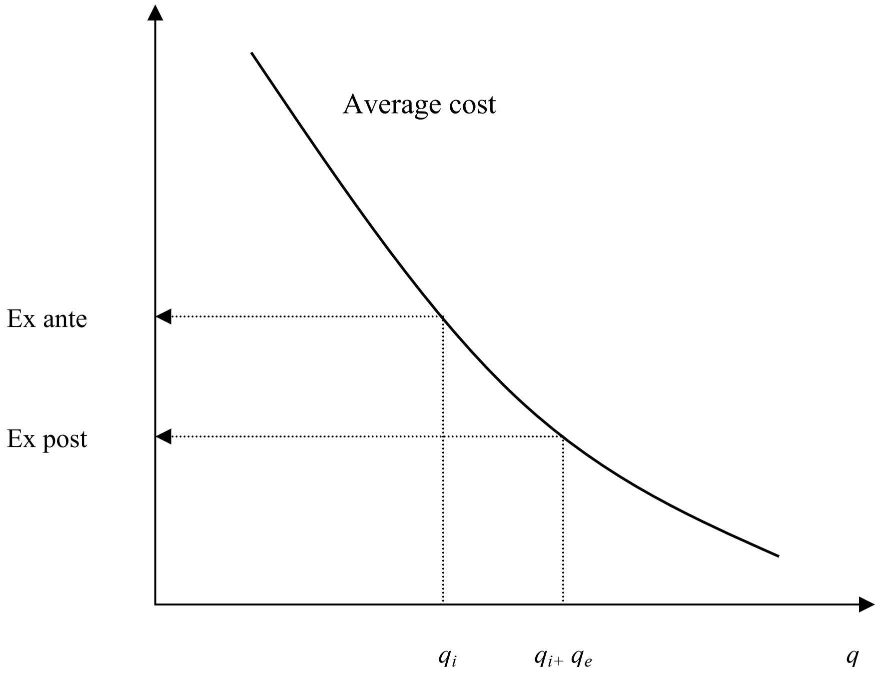


Figure 9: Increasing returns: Opening is beneficial

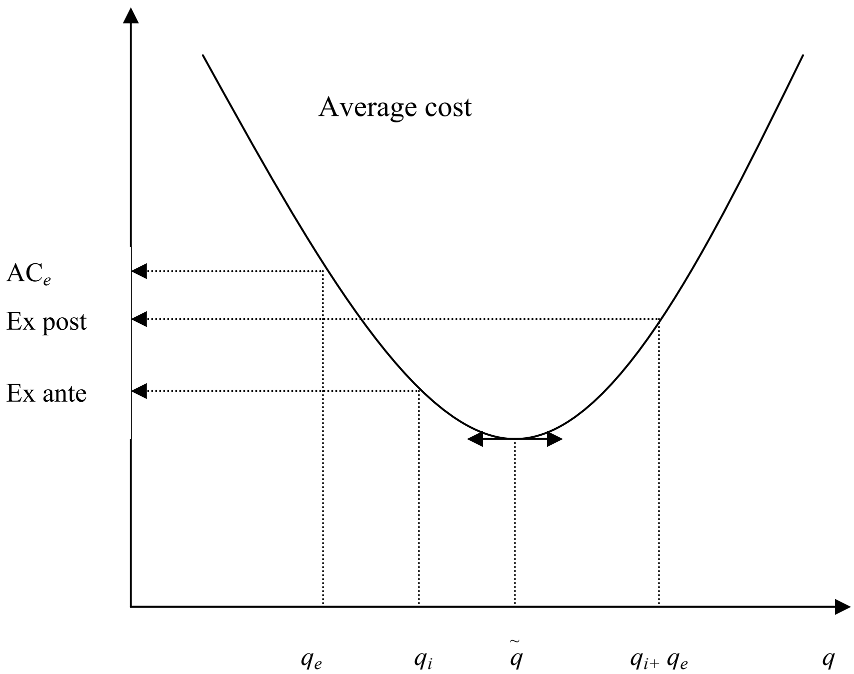


Figure 10: Decreasing returns: Opening benefits only the entrant

c) The promotion of competition should not be a *per se* objective but simply a means to promote efficiency and fairness. In particular, it should not go against the common rules of fair trade, which seek to protect investments of all firms in specific assets, especially intangible assets such as know-how and employee skills. Even when such assets possess the characteristics of a public good (with decreasing average cost of use), it is important to ensure that opening them to competitors does not undermine the incentives for all firms (not just the incumbent) to invest in such assets in the first place).

4. ENTRY AND COMPETITION

4.1. Price competition

Prices are but one of a wide array of business tools that rail firms can use to compete for passengers. In fact competitive strategies concern not just the terms on which a given service is made available to customers, but also the choice of the kinds of service to supply, a choice which has a large number of collateral implications for investment, employment policy, and policy towards acquisitions, outsourcing and joint ventures.

Two features of rail travel make consideration of price competition somewhat different from many other industries. The first is that short-run cross-elasticities between transport modes are rather low, suggesting that for rail to compete purely on price against cars or air travel is not likely to yield rapid profits; at any rate, low price strategies would have to be maintained, and seen to be maintained, over a significant period of years before significant traffic could be gained from other transport modes.

The second is that, because of economies of density, price competition that significantly increases traffic can be an extremely profitable strategy for the firm that undertakes it: the true marginal cost of additional traffic lies some way below the average cost. Thus where on-track competition is feasible, or where the characteristics of a given route suggest inter-modal competition may be unusually keen, the incentives to cut prices can be very strong. This has three important implications. First, stable on-track competition may often not be viable: either it is infeasible, or it is feasible and the result is such fierce price competition that unless the competitors have precisely similar cost structures one of them may be forced to withdraw. This may make it quite difficult to support an industry structure with significant amounts of on-track competition, a fact that should be borne in mind in considering regulatory appraisals of the results of introducing competition. We consider this issue further in section 4.4 below.

Secondly, both entrants and incumbents will seek for ways to soften the impact of competition by differentiating their products. For instance, non-interchangeability of tickets, non-cooperation over scheduling connecting services, different approaches towards discounting and the targeting of different customer groups, may be tempting strategies for all competitors even if their effect on overall customer welfare is negative. Note that this is quite different from similar strategies used with predatory intent, in order to drive competitors out of the market. When there is (successful) predation it is the exiting firm that suffers as well as consumers; when the strategies aim merely at softening competition, the firms benefit and consumers lose.

Thirdly, where inter-modal competition can work (such as on inter-city routes between 200 and 400 km for competition with road and 500km to 1000km for competition with air travel) its effect on prices may be important, and may make price regulation unnecessary in circumstances where it might otherwise have been desirable.

For this to be possible, of course, it is necessary that rail services develop characteristics that make inter-modal competition realistic. High-speed trains have done this with some success (though at high cost) in recent years, and it remains to be considered whether and to what extent other kinds of rail service can provide a significant challenge to other modes, notably the car. We consider this in the next section.

Finally, it is worth noting that developments in communications and information technology, notably of course the spread of the internet, are making an important difference to the sophistication of the pricing strategies that firms can adopt. This is particularly true in the realm of price discrimination. Economists distinguish three types of price discrimination:

i) First-degree price discrimination is charging individual customers according to their own willingness to pay for the good or service. Since this is almost never known to the seller, first-degree price discrimination is typically infeasible, and remains of purely text-book interest.
ii) Second-degree price discrimination, or non-linear pricing, allows customers to select different patterns of payment (usually with a choice between a fixed element and a variable element) in a way that sorts them into groups with different willingness to pay.
iii) Third-degree price discrimination charges different prices to consumers with different observable characteristics, these latter being effectively outside the customer's control, either absolutely (such as age) or for the purposes of the transaction (such as physical location when transport costs are high relative to the value of the good).

Both second-and third-degree price discrimination have long featured in rail pricing, notably through season tickets and discounts for the young and the elderly. However, the internet, and information technology more generally, are making sophisticated second-degree price discrimination easier, notably because customers can be shown, quickly and intuitively, the effect of different pricing packages in a way that allows for an informed choice between them. The effect on third-degree price discrimination is more ambiguous. In some respects such discrimination is becoming harder, because customers can shop around, and it is no longer possible to discriminate between customers according to where they are physically located when they make the transaction. However, this is substantially offset by two other considerations. The first, which is particularly relevant to transport, is that when the product is a service that must be consumed at a certain place and/or time, it remains possible to discriminate between customers according to location or time of consumption rather than location of transaction. Airlines have discovered this in a big way through their computer reservation systems, which can charge very different prices for flights of the same length according to the origin or destination, as well as according to the time of travel[12]. The second reason is that firms can now use sophisticated databases of consumer travel behaviour to target special offers to the individual's presumed preferences. Such targeting has not so far been much exploited in the travel market (unlike in the market for, say, books), but it will not be long in coming.

Overall the increased sophistication of price discrimination is likely to make price competition a more tempting prospect, but also to allow competing firms to segment markets in terms of customer types more effectively than has been possible to date.

4.2. Competition on speed, punctuality and other services

It is evident that the attractiveness of the car lies in its extreme flexibility and therefore its ability to get the traveler to his/her final destination more rapidly than other modes, many of which require changes of carrier or even of mode in the course of a single journey. The dependence of rail on fixed tracks, as well as the economic need to carry passengers in large numbers for average costs to be brought down, makes it implausible that rail could ever reproduce the advantages of the car at all closely. Nevertheless, there are clearly some customers for whom flexibility can be traded off against both speed and price, particularly given the fact that rail can cooperate with other modes. A fall in the price of a train journey allows the passenger more easily to afford a taxi for transit to the final destination. In addition, the growing severity of traffic and parking restrictions in major city centers means that the convenience of the car is diminishing, at least for urban journeys. Similarly, growing security restrictions at airports mean that the advantage of air travel in terms of speed is not as great as

[12] Actually time of transaction (unlike location of transaction) remains a very effective tool of discrimination, because individuals cannot travel freely in time. Tickets booked at the last minute may be very different in price from those booked long in advance.

it was a few years ago. This enlarges significantly the ability of rail travel to approach some of the advantages of its two most important inter-modal competitors. Although some of this development is due to events outside the control of rail companies, their own strategy choices will clearly play an important role here. Some choices are very expensive: on a medium-haul intercity route the investment needed to cut ten minutes from a journey time may be so great that it is not worth making. But other choices – streamlined luggage transfer or portering facilities, smooth connections with buses or with other trains – may make as much different to overall travel times as do improvements in the speed of the trains themselves.

It is worth noting that the same developments in communications and information technology that were noted in the last section make it possible substantially to improve the flexibility of rail services. The container revolution in the freight industry has been of great benefit to the competitivity of rail services. While physical containerization of passengers is not realistically in prospect, virtual containerization may well be. "Total journey" services (see Box 5), in which a traveler uses the internet to book a full through service from origin to destination, using preference parameters stored in a database to choose between multiple modes, may well make rail a central part of future travel demand even if it is rarely the sole mode chosen by the consumer.

Box 4: A "total journey"

I tell the website that I wish to be at DB headquarters in Frankfurt for a meeting at 10 am on the 12th of September, and by default it assumes I shall be leaving from my home (whose parameters it knows). The site has to devise a plan for travel, from the initial alarm call to my mobile telephone, through the taxi reservation and automated check-in for the flight, to the various means of traveling from Frankfurt airport to my destination. It has to ensure that the taxis are waiting and that if there are delays at any stage alternative arrangements are made. If I have not made the journey before, the site may offer me choices of different journey times, levels of comfort, etc., though from my stored parameters it knows how to reduce the set of possible choices to a manageable size. It also knows enough about my preferences to know what complementary services to propose (newspaper, breakfast, haircut…), and enough about my weaknesses to surprise me with the occasional special offer ("Armagnac, sir?"). Though as a traveler I am unaware of this, the site is operated by amazon.com under contract to DB.

One of the unexpected consequences of the privatization and deregulation of the telecoms industry in the 1980s was the vigour with which telecoms operators embraced many dimensions of service competition - telephone handsets stopped being a clunky embarrassment and became a design item, for instance. There is every reason to think that complementary services are likely to become an increasing feature of competition in rail.

4.3. Predation

Naturally the more sophisticated pricing strategies become, and the greater the rewards of effective pricing through the presence of economies of density, the more reasonable it is to fear either that incumbent firms will use predatory pricing, or that firms will claim the presence of predation as a defence against legitimate but vigorous competition. Particularly in network industries, with huge fixed costs and some features of natural monopoly, predation can be a tempting practice for powerful firms (typically though not always incumbents) in order to induce the exit of rivals.

Briefly, the position of the authorities on these matters is as follows. In the USA the standard for decision in these cases (after the Brooke case, resolved by the Supreme Court in 1993) is to prove that the defendant:

- Has priced below some measure of cost;
- Has a good prospect of recouping its losses from predation after the exit from the market of the victim against whom the predation was aimed.

The latter is a much more stringent condition to apply than the former, and has been very difficult to demonstrate in practice.

4.3.1. What is the relevant measure of cost?

There are two relevant measure of cost, for predatory purpose, Average Total Cost (ATC) and Average Variable Cost (AVC); the courts have increasingly being more inclined to apply the second threshold for predation purposes[13]. The choice of the standard can make a big difference in the case of railroads, since the variable cost of an extra passenger can be almost zero. This is why the measure is denominated average variable cost (include the cost of providing the service for all passengers, not just the extra passenger).

In airlines, which is an industry with similar costs structure to railroads, the allegation of predation has been a common feature of the sector[14]. However, to date no carrier has been found guilty of predation.

4.3.2. Department of Transportation (DOT) Guidelines.

These recent guidelines for airlines, issued in 1998, advance a broader interpretation of predation. This practice is deemed not only to relate to price strategies but also to capacity levels. Therefore an incumbent can be blamed for predation if it expands output to a level that makes entry unprofitable for a carrier.

4.3.3. European Union

The European Court of Justice decision on ECS/AKZO chemie BV (in the chemical products industry) case established the doctrine on predation in the EU. AKZO was accused by ECS of pricing very aggressively in the main business line of ECS after the entry of the latter into a new segment of the market.

In 1991, the European Court of Justice resolved that AKZO abused its dominant position by using predatory pricing. The rule was based on the fact that AKZO applied prices below Average Variable Cost (similar to the criteria employed in the USA).

In a more recent decision, (UPS/Deutsche Post 2001)[15] the European Commission found Deutsche Post guilty of using predatory pricing in the segment of parcels distribution. The Commission used the rule of Average Incremental Cost, which is an approximation of Long Run

---

[13] Bolton, Brodley and Riordan (1999).
[14] Peltzman and Winston (2000).
[15] NERA (2001).

Marginal Cost, to judge whether prices were predatory or not. The decision was made on the basis that with current prices the firm could not cover its incremental cost over the following five years.

This decision, in an industry that has some similarities with railroads (in the sense of universal service requirements and a multi-product firm) may be indicative of the criteria and the standard of cost to be applied to future cases in the railroad sector. The AIC could turn out to be a more or less stringent standard for proving predation that AVC (depending on the level of economies of scale), but the former measure, which is more looking forward, has been used also in natural monopoly pricing (especially in telecommunications).

4.4. What is the likely outcome of on-track competition?

Here we try to draw together some of the threads of the preceding discussion by posing the question: what can be expected to happen to market shares under on-track competition, assuming equal access to any infrastructural facilities that have natural monopoly characteristics? In particular, does equal access imply that there are likely to be reasonably equal outcomes as measured by market share? And conversely, if outcomes are not equal will this imply a failure of equal access? The answer to these questions requires us to look at both the demand and the supply side of the industry.

4.4.1. Demand

The reaction of demand depends on the degree of differentiation between the services proposed by the incumbent and the entrants, and on any switching costs that may be incurred when moving from one provider to another. For occasional travelers, there is unlikely to be any switching cost and competition with newcomers will be tough. For frequent travelers, switching costs may be high and the incumbent will probably keep a large market share independently of cost considerations. An additional argument that implies increased switching costs comes from the network characteristics of passenger rail. Only few people would be able to travel point-to-point with a competitor. The majority of customers would have to change trains, partly using local transport as feed, partly using the incumbent.

Let u_i denote the utility of a passenger when travelling in a train operated by the incumbent and let p_i denote the fare. Absent any competitor, the incumbent can charge a price such that $u_i - p_i \geq \overline{u}$ where $\overline{u}$ is the net utility from alternative nodes. When there is an entry, to keep its clients the incumbent has to fix a price such that $u_i - p_i \geq u_e - p_e$ where u_e and p_e are respectively the utility and the price of the service provided by the entrants (and assuming that entry is feasible, which means $u_e - p_e > \overline{u}$). The difference $u_i - u_e$ is the value of the incumbent's advantages: the higher the switching costs, the larger this utility differential. For occasional travelers, $u_i = u_e$ so that $p_i \leq p_e$ is necessary to keep these clients, which is feasible only if the incumbent has a cost advantage, that is when $c_i < c_e$.For frequent travelers, a tariff such that $p_i \leq u_i - u_e + c_e$ prevents any entry. Even if $c_i > c_e$, the switching cost $u_i - u_e > 0$ allows the incumbent to cut the entrant 's price.

Switching costs (natural or strategic) are a strong limitation to competition, as has been well documented in banking, telecommunications, electricity distribution, air transport and pharmaceuticals. As a consequence, incumbent firms will probably keep dominant market shares in subsets of weakly flexible demand, and will lose shares in relatively contestable sub-markets (this may also be a consequence of natural utility advantages of remaining with incumbent suppliers).

4.4.2. Supply

The nature of competition will also be determined by the nature of the costs faced by the entrant and the incumbent: this will affect how low a price each can afford to set in order to attract customers. Other things equal, the lower is the entrant's cost of operations relative to that of the incumbent, the more intense will be the nature of price competition and therefore the higher the likely market share that the entrant can attract.

However, when the incumbent operates a network and the entrant competes only on point-to-point routes, there is an important source of asymmetry induced by network effects. For the entrant, the marginal cost of an additional train full of passengers is the cost of running the train (including administrative costs), plus the access charge for the service. For the incumbent, the true marginal cost consists of the same elements as for the entrant[16], *plus* an additional element (the *opportunity cost*), which is any additional net cost incurred on those connecting routes to which some of the passengers may subsequently transfer[17]. When transferring passengers in fact yield a profit on the connecting routes, this opportunity cost will be negative, and an incumbent's true cost will lie below a conventionally-measured accounting measure of its costs of providing the service. Three consequences follow from this:

i) First, in networks where connecting traffic is a comparatively large fraction of overall traffic (like Germany but unlike France, say), there will be fewer cherries for entrants to pick, i.e., there are relatively few connections with a high point-to-point demand.
ii) Secondly, even when entrants appear to enjoy a cost advantage as normally measured, the opportunity-cost element will mean that the incumbent is a tougher competitor than this advantage would indicate (because it has an incentive to protect its connecting traffic). Where entrants do compete head-to head with incumbents their likely market share will be lower than conventional cost comparisons would lead us to predict.
iii) Thirdly, it is likely that the Mohring effect (see Box 2 above) means that the opportunity cost element will become more important as the entrant's market share increases, since the reduction in the value of frequency of service to passengers becomes progressively more important as the frequency itself declines. The cancellation of half the services is more costly to passengers if services previously ran every two hours than if they ran every fifteen minutes. This implies (see Figure 11) that not only will the incumbent's true cost lie below its accounting cost, but it will be significantly more steeply sloped. Therefore even an entrant with a significant initial cost advantage will find that as it eats into the incumbent's market share its cost advantage is progressively eroded. To put it another way, the eventual market share of the entrant is likely to be less sensitive to its initial cost advantage than if opportunity cost considerations did not play a role (in the figure equilibrium market shares are drawn where marginal costs of incumbent and entrant are equal). Without opportunity cost considerations, an entrant with an initial cost advantage could easily reach a large market share, but when opportunity costs matter its market share will be unlikely to become very large.

[16] These same elements may of course have higher values for the incumbent, for instance if the incumbent faces diseconomies of complexity from running the network.
[17] This is similar in spirit to the opportunity cost calculation that underlies the Efficient Component Pricing Rule for access price regulation (sometimes called the Baumol-Willig rule). The difference here is that we are considering a complementarity between two services on both of which there is competition, rather than a complementarity between shared infrastructure with mandated access and a competitive downstream service.

(This argument assumes that an entrant's services are not perceived by the passenger as contributing to the overall frequency of the service, perhaps because of non-transferability of tickets.).

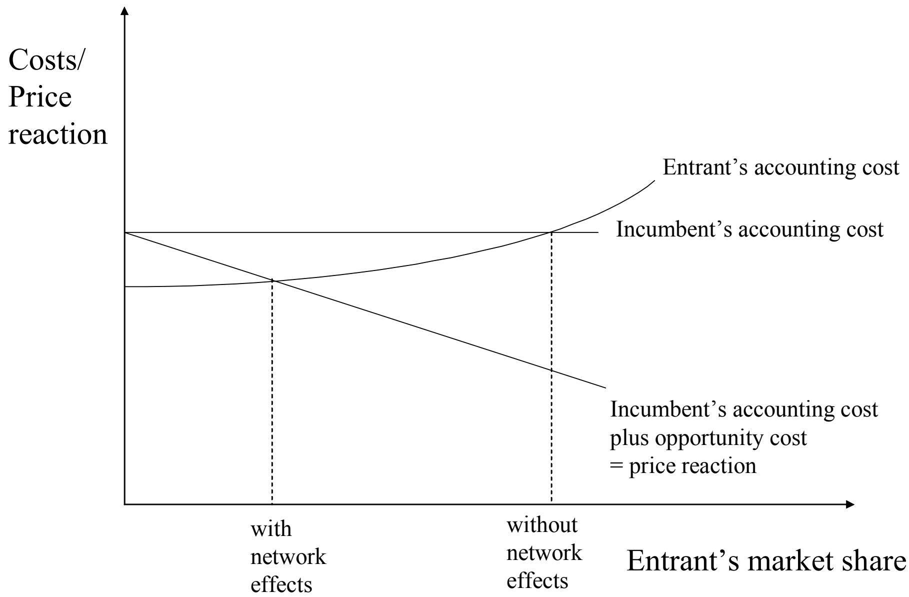


Figure 11: Marginal passenger costs at different levels of entrant market share

The overall conclusions to be drawn from this line of reasoning are twofold. First, given the unavoidable asymmetry between a network operator and an entrant on point-to-point routes, it may be unlikely that effective competition will lead to large market shares for the entrant. This is not, however, to say that we can predict exactly how large such market shares will be, since circumstances will vary significantly from route to route. Secondly, the share of the incumbent will be larger the more polycentric the network and resulting effects are. Thirdly, this asymmetry is not a sign of a market regime failure. Fourthly, the opportunity costs of network traffic are genuine social costs, so that considering the "success" of competition purely in terms of the market shares gained by entrants on point-to-point routes would be seriously short-sighted. If these market shares come at the expense of disruptions in network connections they may well be symptoms of the failure of competition rather than its success.

Finally, an important point to note is that if the entrant has a significant cost advantage over the incumbent in providing the service, then the incumbent may have an interest in arranging interconnections so as to capture as much of the network traffic as it can. When the costs of interconnection can be avoided, so that passengers can easily switch operators to make through journeys, then the incumbent's connecting services become complementary to the point-to-point services of the entrant. In these circumstances the incumbent may not only not be damaged, but may positively benefit from the cheap fares provided by the entrant, since these will increase demand for the through journey.

This leads naturally to the question how to judge the value of interconnections, which we take up in section 4.5.

4.5. The value of interconnections and its relevance for horizontal integration decisions

The advantages of joint production as compared with separated production are typically considered to arise in the presence of "economies of scope". Specifically, if a quantity q_a of product a and a quantity q_b of product b are to be produced, it is less costly to produce them within the same firm, i.e.,

$$C(q_a, q_b) < C(q_a, 0) + C(0, q_b).$$

These economies of scope are explained in terms of

- better organization and coordination;
- simultaneous use of "public" inputs (e.g. data bases).

It is certainly true, on the other hand, that too much diversification can impair the development of a firm and create some social inefficiencies. All synergies have their limits - as shown by the recent failure of the merger between AOL and Time Warner. Nevertheless, even if $C(q_a, q_b) > C(q_a, 0) + C(0, q_b)$ is a necessary condition for separation to be desirable, it is by no means a sufficient condition.

The first reason is the way costs are evaluated in networks. Suppose an elementary transport network made of two segments as in Figure 12.

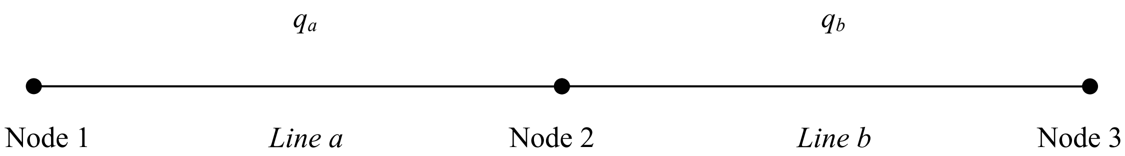


Figure 12: Two-segment transportation network

The infrastructure is made of three nodes and two lines, and there are six products:

- transport from 1 to 2 and back;
- transport from 2 to 3 and back;
- transport from 1 to 3 and back.

Suppose that the infrastructure cost is clearly evaluated and a fair fee is to be calculated for an entrant who wants to compete with the incumbent on line a. What is the accounting cost of using only line a? Does it include all the costs of node 2 or only the costs of arriving at node 2? Are the costs of connecting arrival (from node 1) equipment and departure (to node 3) equipment to be included or not? If these interconnection costs (which include not just physical equipment costs but planning and organizational costs) are not included in the accounting cost of operating line a, there is no reason to include them in the accounting cost of operating line b. Consequently, they are taken into account only if a firm provides the whole trip from 1 to 3 and, for this reason, even when there are economies of scope it can easily appear that

$$C(q_a, q_b) > C(q_a, 0) + C(0, q_b).$$

So, we see that accounting rules badly adapted to the specificity of networks can result in a false interpretation in terms of efficiency.

The second reason is that efficiency on the supply side is not the only relevant consideration. What about travellers' utility? Interconnection, compatibility, "one-click" registering, etc. are highly valued by passengers. These extra services plus the indirect externalities due to the installation of interconnectors[18] have the consequence that utility functions are super-additive:

$$U(q_a, q_b) > U(q_a, 0) + U(0, q_b).$$

Indeed, passengers are less tired and stressed when they can go through one single operator for a two-stage trip than if they have to wait and register two times, to handle their bagages at the intermediary node, etc.

For this reason, the true test for disintegration should be:

$$U(q_a, q_b) - C(q_a, q_b) \gtrless \left[U(q_a, 0) - C(q_a, 0)\right] + \left[U(0, q_b) - C(0, q_b)\right].$$

When the net utility from an integrated structure is higher than the net utility from a disintegrated structure, disentangling would be socially wasteful.

To a certain extent, on-track competition can allow the preferences of travellers to be revealed directly, since they may face a direct choice between a (possibly more expensive) through service and a (possibly cheaper) service that requires changing operators. But once again, this suggests a reason why on-track competition may lead to relatively low market shares for entrants, even when it is working well.

[18] The more numerous the passengers who use the interconnectors are, the higher are the incentives for the operator to develop them and to upgrade their quality.

5. MEASURING PERFORMANCE IN THE INDUSTRY

5.1. Ideal measures of performance

Most economic analysis would begin from the assumption that the appropriate and theoretically rigorous measure of performance by which to judge railway reforms would be some measure of consumer and producer surplus, with room for argument over the appropriate weight to be placed on these two components. In principle, demand analysis should allow for the estimation of changes in consumer surplus after a change in prices and service quantities and other characteristics. There are certain well-known issues that arise in such contexts (such as index number problems, including the choice between so-called compensating and equivalent variation measures of welfare changes), but in principle the methodological issues as reasonably clear-cut. Such demand analysis essentially computes a money value for the changes in service characteristics, using data on past consumer choices to reveal how much consumers appear to have valued the characteristics in question. Such methods have been used in ex post analyses of reform in a number of industries (see Morrison and Winston, 1986, for the airline industry).

In practice, however, data are often not available in sufficient detail to compute measures of surplus. And when they are available they are often available only with a lag, which means that policy-making, by regulators, politicians, or executives of the firms themselves, typically requires measures of performance that are rapidly available and intuitively comprehensible. Such measures typically fall into two main groups, as section 5.2 now describes.

5.2. Practical measures of performance

Practical evaluation of performance in the passenger transport industry typically uses one of two approaches:

1) Various measures can be calculated of characteristics of the industry (output per person, profits, punctuality, reduction of public subsidy, for instance), and these can be presented without any explicit effort being made to aggregate or compare them. This then allows for a debate to occur as to the appropriate weight to be placed on any of these proposed measures.
2) Measures can be aggregated into indices that are argued on theoretical grounds to capture the appropriate weights of different components in some suitable measure of social welfare. Often such aggregation is partial (total factor productivity indices, for instance, measure technical efficiency but not allocative efficiency).

Table 16 presents some measures of the first kind, drawn from the European SORT IT project which has analysed data made available by the UIC (Union Internationale des Chemins de Fer). Note that the characterisation of the firms as state-controlled or commercial is due to SORT-IT and may not capture the most important governance characteristics.

Table 16: Comparison of European Rail Operators (1994 and 1997)

State-Controlled Firms	Operating Performance		Commercial Performance		Financial Performance	
	Vehicle Kms/ Number of Staff		Traffic Units/ Vehicle Kms		Total Revenue/ Total Cost	
	1994	*1997*	*1994*	*1997*	*1994*	*1997*
VR Finland	2.540		301	(304)	0.87	(0.81)
SNCF France	2.747	(3.059)	224	(225)	0.50	(0.44)
DB Germany	2.694	(3.120)	150	(158)	0.40**	(0.74)
CH Greece	1.060	(3.593)	144	(119)	0.17	(0.13)
CP Portugal	2.449	(1.722)	195	(140)	0.37	(0.38)
RENFE Spain	3.746	(3.711)	151	(167)	0.36	(0.44)
CFL Luxem.	2.416	(4.536) (2.539)	129	(121)	0.29	na
Mean	2.522	(298)*	185	(23.1)*	0.42	(0.84)****
Commercial Firms						
	1994	*1997*	*1994*	*1997*	*1994*	*1997*
OBB Austria	2.170	(2.505)	163	(167)	0.38	(0.39)
SNCB Belgium	2.355	(2.560)	163	(159)	0.21	(0.30)
DSB Denmark	3.866	na	132	(112)	0.45	(0.89)
CIE Ireland	2.773	na	134	(89)	0.79	na
FS Italy	2.256	(2.876)	222	(210)	0.44	(0.29)
NS Netherland	4.435	(4.674)	147	(147)	0.54	(0.41)
NSB Norway	3.862	(2.580)	137	(150)	0.39	(0.75)
SJ Sweden	4.926	(8.990)	252	(237)	0.42	(0.51)
CFFSwitzerl.	3.516	(3.758)	165	(175)	0.46	(0.44)
BR U.K.	3.017	na	120	na	0.74***	na
Mean	3.318	(302)*	164	(13.3)*	0.48	(0.054)***

Notes:
* Standard deviation of the mean. na – not available.
** The DB figure is for 1993, since the financial statistics for the newly merged DB AG appear to be inconsistent with previous years.
*** The BR figure is for 1993, since the statistics didn't take into account the huge increases in track access charges levied by Railtrack in 1994.
Source: Shires (1998).

However, closer investigation reveals that the distinction between these two types of measure is less sharp than it first appears. The measures in Table 16 are already aggregates to some degree – for instance, "traffic units" is the arithmetic sum of passenger-kilometres and tonne-kilometres using an arbitrary one-to-one weighting. In more explicit aggregation exercises[19], we would expect the index of aggregation of different outputs to be weighted by the revenue share of each product. Such a measure would show more passenger-dedicated firms to have a higher commercial performance, since the average price of a passenger-km is between 40 to 419 times the average price of a ton-km (see Table 2). So, the index reported in the second column of Table 16 has an important distortion.

[19] Oum, Waters and Yu (1999).

Table 17: Sources of distortion in aggregation of traffic units

Country	Passenger-km as a percentage of total traffic units	Ratio of average passenger tariff over average freight rate[20]
Sweden	24	419
Austria	40	103
Germany	47	224
Belgium	48	119
France	55	126
Spain	61	121
Ireland	69	157
Italy	69	157
Portugal	72	82
Netherlands	84	169
Greece	84	40

Source: See Thompson (1997).

A second reason for caution is that factor productivity indices in network industries are highly dependent on exogenous network characteristics (such as length of track or network configuration). Scale and density factors have been highly significant as we noted in section 2.3. Therefore, the comparison of gross productivity measures like those in Table 16 omits important characteristics that should not be overlooked in assessing the performance of the firms in question.

We therefore present some measures of the second kind, namely explicit aggregate indices. The report of the Australian Productivity Commission published indicators of total factor productivity (TFP) for combined freight and passenger services for rail systems in Europe, Australia, New Zealand, USA, Canada and Japan. TFP measures the global productivity of firms taking in account all the inputs and outputs. The methodology employed is DEA -Data Envelope Analysis (Charnes *et al*., 1978)- which ranks the productivity of the different companies respect to the best practice case (to whom is given the value = 1.0). The indicators of productivity calculated are **Global Productivity** (which takes into account the entire factors affecting productivity) and **Technical Efficiency** (TFP adjusted for an output size effect). As we can observe in table 18, the values and the position in the ranking change when we change the estimator.

[20] Average rate = passenger (freight)revenue/ passenger (Ton) - km

Table 18: Global productivity and technical efficiency measures for European operators

Rank		Global Productivity	Rank		Technical Efficiency
1	Netherlands	0.52	1	Ireland	1.00
2	Ireland	0.41	2	Luxembourg	1.00
3	Sweden	0.38	3	Norway	0.74
4	France	0.38	4	Netherlands	0.68
5	Spain	0.38	5	Portugal	0.63
6	Finland	0.35	6	Denmark	0.60
7	U.K. (94)	0.34	7	Sweden	0.47
8	Portugal	0.33	8	Finland	0.45
9	Italy	0.32	9	Spain	0.43
10	Switzerland	0.32	10	U.K.	0.43
11	Germany	0.31	11	France	0.39
12	Norway	0.31	12	Switzerland	0.38
13	Denmark (95)	0.28	13	Italy	0.33
14	Austria	0.25	14	Germany	0.32
15	Belgium	0.20	15	Austria	0.27
16	Luxembourg	0.20	16	Belgium	0.24

Notes: Productivity Measure for Combined Passenger and Freight Services Using Data Envelopment Analysis Methodology (The base year is 1997 unless otherwise stated).
Source: Productivity Commission Report 1999, Australia

Table 19, by contrast, presents results of an explicit econometric estimation of costs of rail operation, drawn from Cantos. These differ from those in Table 18 in two main ways. First, they use cost-minimization rather than output maximization (given inputs) as a criterion of technical efficiency, and although one implies the other when all firms face the same parametric prices, this last assumption may not characterize the effective environment for the rail operators under investigation. Secondly, they use parametric regression techniques which have the disadvantage of requiring explicit functional forms but the advantage of being less vulnerable to errors and omitted heterogeneity in the data.

The table shows by how much the cost of various operators exceed those of the SNCF once different output characteristics have been controlled for. The data are a panel of 12 countries and 17 years (1974-1990). All the values are statistically significant at 5%.

Table 19: Intercepts of cost function for different European operators

Country	Dummy Estimation	Standard Error	t-statistic
DB	0.522	0.099	5.273
SJ-BJ	1.064	0.522	2.038
RENFE	1.251	0.468	2.673
BR	1.263	0.339	3.726
FS	1.529	0.371	4.121
NSB	1.666	0.742	2.245
VR	1.709	0.684	2.499
NS	1.970	0.758	2.599
DSB	1.997	0.781	2.557
OBB	2.263	0.673	3.363
SNCB	2.302	0.738	3.119

Source: Cantos (2001).

In principle, under reasonably competitive conditions, productivity measures (whether output or cost-based) should give a fairly reliable indication of likely benefits to consumers. However, when there is monopoly power, productivity gains may not be passed on to consumers but be taken as profit by the shareholders of the firms. Conversely, when there are public subsidies, variations in the level of public support may have an impact on consumer welfare independently of the evolution of productivity.

We now consider briefly the evidence linking performance to competition.

5.3. How is performance related to competition?

Identifying the possible influences of competition on firm performance is not easy. Even if the degree of competition it faces has no direct causal influence on the behaviour of any individual firm, it may be that more competitive market environments see a faster replacement of relatively inefficient by relatively efficient firms. In this case, a correlation emerges over time between a measure of competition at industry level and the average efficiency of those firms that survive. Even if survival as such is not differentially affected, the degree of competition may affect how large a share of output is occupied by the products of relatively efficient firms. In short, competition may work not just through *incentives* but also through *selection* (see Carlin, Haskel and Seabright, 2001).

In fact it is quite likely that competition does have a direct influence on behaviour via incentives, but economic models show that the effect may be ambiguous. One example of ambiguity comes from Willig's (1987) model, in which he demonstrates two offsetting effects of increased competition on the incentives for managers to exert effort. Whilst increased competition makes profits more sensitive to managerial effort, it also depresses demand for the firm's output, which dampens profits and hence blunts the incentive.

In the innovation literature, there are models that suggest that more competition is good for innovation and others that highlight a hump-shaped relationship, in which a moderate degree of competition is better than either monopoly or intense competition. For reasons first suggested by Schumpeter and recently analyzed more formally by others (see Aghion and Howitt, 1998, for example), some degree of prior market power may be important in providing firms with sufficient retained earnings to finance investment. Moreover, the prospect of some future profits may be essential to ensure that current retained earnings are indeed invested instead of wasted. Other variants stress a monotonic relationship with greater competition inducing productivity growth. For example, the emergence of new competitors threatens the temporary monopoly profits from innovation and increases the incentive of the incumbents to shorten the innovation cycle (Aghion, Dewatripont and Rey, 1997). However, many models stress that at least some degree of competition is necessary to enable managers to be given adequate performance incentives (see Meyer and Vickers, 1997), though there may be one or two particular industries where the natural monopoly characteristics of the technology override incentive considerations. The controversy arises mainly over whether competition continues to be beneficial for incentives once a small number of competitors are already present.

Empirical support for the role of competition as a spur to performance comes from recent econometric research using a variety of performance measures. For instance, Blundell *et al*. (1995) use numbers of innovations as a measure. The results are consistent with those of a quite different methodology (bench-marking using case studies) in which Baily and Gersbach (1995) found that "head-to-head" competition in the same market resulted in faster innovation in several manufacturing industries. Nickell (1996) controls for industry level concentration and import concentration and tests whether a firm-level measure of competition is correlated with performance. He finds that indicators of competitive pressure at firm level are significantly related to the level and growth of total factor productivity. In an empirical study of entry thresholds, Bresnahan and Reiss (1991) found that most of the competitive impact from entry comes from the first two entrants to challenge a monopolist, with the effect levelling out once the number of market participants is around five. Other results come from industry studies (Kwoka, 1996; Neven and Roeller, 1996; Ng and Seabright, 2001), from cross-sectoral surveys (Carlin *et al*., 2002) or from economy-wide studies using indices to proxy for competitive effects across a range of sectors (Nicoletti and Scarpetta, 2002).

6. REGULATION

6.1. Different regimes around the world

The table in Appendix 6 provides a comprehensive description of regulatory regimes in a large number of countries. The main important dimensions along which regimes differ are the following:

i) **Ownership**: Are operators privately or publicly owned (or both)?
ii) **Integration**: Are operators vertically and/or horizontally disintegrated?
iii) **Price regulation**: Are operators free to set some or all fares?
iv) **Yardstick regulation**: to the extent that operators are regulated, is comparison between operators an explicit component of the regulatory contract?
v) **Competition for the market**: Do operators compete for the exclusive right to operate services on infrastructure for a certain period of time?
vi) **Competition in the market**: Do operators compete by sharing tracks?

Broadly speaking five different kinds of regime appear to have evolved:

i) A regime with privatization, vertical integration, regional disintegration and yardstick regulation (seen in Japan, where inter-modal competition is weak).
ii) A regime with privatization, vertical integration, regional disintegration and no regulation (seen in the USA, Australia and New Zealand, in all of which inter-modal competition is strong).
iii) A regime with privatization, vertical disintegration, regional disintegration, price regulation and competition for the market (the UK).
iv) A regime with continued state ownership (to date), vertical and regional integration, no price regulation, the introduction of competition in the market for inter-city travel, and competition for the market in suburban travel (Sweden and Germany).
v) A regime with continued state ownership (to date), vertical and regional integration, no price regulation but no attempts to introduce competition (France and Spain).

It will be important to consider to what extent the different regimes represent alternative visions of the nature of rail transport, rather than adaptations to the different circumstances of each country.

6.2. Messages from the case studies

Bearing in mind the principles of competition in rail transport that were outlined earlier in this document, it seems reasonable to draw the following conclusions from the case studies:

i) The comparatively light regulation seen in the United States, Australia and New Zealand appears to be due to the intensity of inter-modal competition. Rail has very small market share, and consequently no temptation for commercially-run railway operators simply to exploit short-run monopoly power.
ii) The apparent effectiveness of rail reorganization in Japan is due in no small measure to the use of strict yardstick regulation of fare-setting, made necessary by the weak degree of inter-modal competition in Japan and the consequent temptations of monopoly power.

iii) In Europe where inter-modal competition is intermediate in strength, it is doubtful whether parallel monopoly with vertical integration as in Australia would be enough to ensure successful rail performance, since most estimates indicate that inter-modal demand elasticities are small, at least in the short run. However, significant uncertainty remains about the appropriate elements of a suitable regulatory framework. The alternatives to Japanese yardstick regulation appear to involve either partial vertical disintegration with on-track competition, or complete vertical disintegration with contractual franchising (otherwise known as "competition for the market").

iv) On-track competition may have a significant effect on increasing the incentives of incumbents to cut costs and improve services, though it is highly unlikely that entrants will succeed in establishing large market shares. Entrants will tend to be low-cost operators (necessarily so because of their lack of network advantages). As the German experiences have indicated, their presence will be seen on certain routes with high point-to-point traffic, not over the whole of the network. Nor, in view of the value of network connectivity, is it particularly desirable that it should.

v) The UK experience illustrates the high importance of a clear and consistent regulatory framework, with a single regulatory responsibility and avoidance of frequent changes in the rules. The question whether it also indicates fundamental flaws in the franchising system is less clear since so many other changes have been simultaneously implemented.

vi) The long-run coordination of investment and service provision is also essential if competition in railways is not to lead to a collapse in infrastructure investment (as in the UK example). This does not in itself imply that vertical disintegration should be avoided, but that mechanisms for achieving coordination (including possible long-term contracts) may need to be put in place.

The question of whether a tendering system (as in the UK) is preferable to a system of open access or on-track competition (as in Germany) is difficult to answer on the basis of case-study evidence. The one country where tendering has been tried systematically for inter-city traffic (the UK) is also one where so many other regulatory changes took place at the same time that it is hard to judge whether the evident failures of the experiment (including the bankruptcy of railtrack) reveal anything about the merits or defect of tendering at all. Nevertheless, on the basis of some general considerations we can set out one or two likely advantages of tendering and open access, even if the precise magnitude of these advantages awaits further experience and research.

6.2.1. Advantages of tendering

i) Long term contracts are possible between the infrastructure operator and the service operators (up to the limit of the franchise length), including investment commitments.

ii) Franchise contracts can specify characteristics of services (frequency, punctuality targets etc) – but: policy failure versus market failure.

iii) Franchisees are able to provide services in whole geographical regions and thereby internalize more of the relevant network externalities than are open-access entrants on point-to-point services.

6.2.2. Advantages of open access

i) A single network operator can optimize network functioning as a whole in order to internalize the externalities due to flows of connecting traffic.

ii) Competition is not influenced by arbitrary boundaries between franchise areas.

iii) Competition on the same track has a much more direct and immediate constraining effect upon market power than the prospect of losing a franchise at some point in the future.
iv) Planning of services is based on entrepreneurial decisions with flexibility of entry and exit.

Nevertheless, it should be stressed that the precise degree to which these advantages matter in practice depends on the specificities of the case in ways that we still need to understand.

7. CONCLUDING REMARKS

This survey has considered and summarized a large volume of both theoretical and empirical material on the nature of competition in rail transport. We conclude by emphasizing two key points.

First, railways are not like any other industry, not even any other network industry. Their technology has changed much more slowly than in telecoms, they face much greater economies of density than electricity, their investment and operations decisions require much closer co-ordination than in air transport, the coordination of different operations is significantly more complex than in other networks, and they have more substitutes in the lives of consumers than the services of other network industries.

Secondly, and in consequence, the fact that we still have limited evidence about the results of different regulatory reforms in the industry should make us double cautious. It is too early to be sure what works best – and mechanically applying solutions that appear to have worked elsewhere may be very ill-advised. The next few years will doubtless provide much fascinating evidence and experience to guide both policy and research.

APPENDIX 1: THE SWEDISH CASE

The restructuring of Swedish Railroads began with the 1988 Transportation Act. The motivation of the reform was to increase productivity and address the lack of investment in the service. Additionally, the integrated publicly-owned monopolist was losing a lot of money.

The New Industry Structure

The new configuration of the market is characterized by the **vertical separation** between infrastructure provision and train services, and also by the **horizontal disaggregation** of the train business in trunk system (intercity services) and county areas (local services).

Vertical Separation

A state agency, Banverket (BV), owns the infrastructure and has the responsibility for investment and maintenance of the track as well as for service scheduling and safety regulation. This firm charges train companies an access fee equal to the marginal social cost of usage. There is no goal of full recovery cost of infrastructure. This policy, which is also used for road tolls in Sweden, is close to first best pricing, though not exactly equal to it since marginal social cost does not cover future infrastructure investment.

The main train operator company (SJ) runs passenger and freight services. It has the exclusivity of service in the trunk system and is also present in the county areas. This company, though state owned, has to operate under commercial basis, receiving no subsides from the state.

Horizontal Disaggregation

The trunk system corresponds to the national network, which is served by the incumbent SJ in the profitable segments under exclusivity with no regulation of prices. In the non-profitable routes, the service is allocated through tendering process under lower subsidy criteria.

The county areas are under the jurisdiction of the local authority. They franchise the service through competitive bids. In some cases local authorities own the rolling stock and franchise the operation and maintenance. This reduces the entry barriers to potential providers different from SJ. However, only few cases of successful entry have occurred.

Competition

At the time being, competition has occurred through franchising in the local services and non-profitable trunk routes. For the rest of the trunk system, some form of competition is expected to occur in the near future.

The experience has not led to much entry, with the incumbent operator winning almost all the franchises. One entrant, BK-Tag, has managed to obtain a contract in two cases. This may reveal the advantage of SJ in terms of economies of scope and scale, but also it may raise suspicions about anti-competitive behavior from the part of the main operator. In 1993, BK- Tag complained of a predatory offer from the part of the incumbent in a franchise. The Swedish Competition Authority considered that the incumbent had abused its dominant position in the industry by making a predatory offer (under variable cost), thereby infringing article 19 of the Swedish Competition Act.

The decision was intended to be based on the criteria employed in the AKZO case (AKZO Chemie BV vs Commision July 3, 1991), where the price is deemed predatory if it is below average variable cost. Since it was not possible to define which costs were fixed and which variable, finally the judgment was based in the fact that the bid was too low to cover all the cost of providing the service.

Quoting OECD (1998):

"The authority concluded that the Swedish State Railways – due to its dominant position in the market with greater financial resources including the posibility of cross subsidiation and its previous legal monopoly – has a strong responsibility not to behave in a way detrimental to the competition on the market and thereby preventing potential competitors from entering this market".

The outcome of the investigation was an injunction against SJ with damages claim of 30 Million Kronas (3 Million USD) for using predatory pricing. The affected firm appealed to the national court.

Outcomes of the rail reform

Decrease in Subsidies

In the non-profitable inter-regional lines, allocated through competitive franchises, the subsidies required to operate have decreased by 20% to 30% with respect to the previous level (OECD 1998). Franchises at the local level have also produced a decrease in subsidies of between 20% and 40% (Shires *et al.* 1999a).

Commercial and Operational Performance

SJ, the main train operating company, turned the operating loss of US$ 122 million in 1988 into a profit of US$ 72 million in 1995 (Anders Lundberg, 1996 JRTR). However, it should be borne in mind that in 1988 SJ was in charge of passenger and infrastructure operations, while in 1995 it ran only passenger and freight services.

The SORT IT study (Shires *et al*. 1999b)., using a non-parametric index analysis, reported that operational costs had been reduced by 10% after vertical separation in Sweden. Fares declined from 11 US cents per passenger km in 1990 to 7.9 cents in 1999.

Market share

Table A1 shows that Swedish railways have gained market share in the 1990s, in contrast to the European average case where rail continued to lose market share.

For intercity travel (more than 100 km) railroads had a share of 13% in 2001 of which the incumbent, SJ, had 90% (SJ Annual Report 2001). In 1995 the rail market share was only 10% (OECD1998).

After cars, railroads have the largest market share of domestic travel, with close competition from airlines. The latter mode dominates for trips over 500 km. On the busy route Stockolm – Gothenburg the market split is almost 50-50.

Table A1: Rail Passenger Traffic (Billion of pkm) in Sweden and Rail Market Share

		1990	1994	1995	1996	1997	1998	1999
Traffic	$10^{E}9$ pkm	6.0	6.1	6.4	6.4	6.4	7.1	7.4
Sweden	%	5.2	5.8	5.8	5.8	5.8	6.2	6.4
Europe	%	6.7	6.2	6.1	6.2	6.1	6.0	6.1

Source: Union Internationale des Chemins de Fer (2002).

Conclusion

The restructuring of Swedish railroads appears successful when considered in terms of better productivity, lower cost and increasing market share achieved in the post reform period; though a rigorous ex post evaluation has yet to be conduted. The new structure of the industry has also accomplished a better allocation of public subsidies in the sector. This is a case where the success comes not from the change of ownership (privatization), but from a good design of market structure with clear objectives and incentives for the firms being created.

It is not so clear, though, that the result has much to do with introducing competition in the traditional sense of the term. Although in long distance routes, railroads face non-regulated competition from other modes (bus, coach and airlines), within the rail sector there have been no clear challenges to the incumbent. If the authorities want to encourage competition, some actions like extending franchise periods or leasing rolling stock to entrants may need to be undertaken. Other more drastic measures may include restricting the incumbent from participating in licenses or the division of the incumbent into two or more train operating companies. However it is not clear that the existence of several providers in different routes will guarantee a better outcome in terms of prices and quality of service.

APPENDIX 2: THE JAPANESE CASE

In 1987, the Japanese Railroads JNR was drastically reorganized. The nationwide, state owned enterprise was split into six regional passenger firms and one national freight company. Over the subsequent years a process of privatization was begun. It has still not been completed, though the last remaining government stake in the largest company, East Japan Railways, was sold in the summer of 2002. The main goal pursued with this measure, as in other countries, was to reduce the state subsidies in the sector as well as to improve the productivity and efficiency of the service.

In Japan, railroads have always played a relevant role in the passenger transportation sector (27.3% of the travel share in 1998)[21] mainly due to the high population density of the country. Before the privatization, JNR already competed with several private railroads in urban travel and with airlines for medium distance travel (high-speed lines).

The key aspects of the Japanese model

Horizontal Separation

A nation wide company was considered too big for being efficiently managed. Six regional firms were created, in basis of geographical as well as cost-minimizing and whole-network self-financing criteria. With this subdivision the 95% of the trips are completed inside the network of each company.

Vertical Integration

The new firms own the track and provides the passenger transportation service. This allows economies of coordination between both activities and also entry into non-rail business (housing development, shopping, etc) which may allow internalization of the positive externalities created by the train service. The option of on-track competition was not considered as the best one since regional firms face already competition from private railways, which have their own track, and also from airlines for medium distance travel.

Functional Separation

Freight was considered a separate business with respect to passenger transportation. Freight companies utilize the regional firms' track by paying marginal cost of usage.

Yardstick Competition

Although regional firms do not compete with each other in their respective areas, there is indirect competition through performance. Using the concept of yardstick competition the regulator uses performance indicators of the firms in order to set the appropriate rates. As it is shown below, since privatization, average fares have decreased for four of the firms.

In the spirit of providing more management autonomy to new firms, the regulatory burden over firms was alleviated. Rail firms now count on more flexibility for setting investments, itineraries, frequencies, and other business decisions.

Evaluation of the reform

Based on the goals proposed by the authorities, the outcome of the restructuring can be considered as a success. Consumers have gained in terms of lower prices, better quality and safety. Rail firms have been able to match the increasing demand in terms of capacity and product

[21] For Europe, the share of rail transport is around 7%

differentiation. The state has reduced the transfers and firms have also improved productivity, which has allowed them to have operating profits and also to reduce their heavy inherited debt levels.

Table A2: Performance change between 1987 and 1998 (Values correspond to percentage change respect to 1987)

Item	Large Private Operator	JR East	JR Central	JR West	JR Hokkaio	JR Shikoku	JR Kyushu
Average Fare	29.3	-8.4	-2.0	-7.9	-3.1	6.1	2.6
Demand	1.8	20.7	18.0	16.9	15.8	8.5	8.0
Labor Productivity.	27.7	52.7	12.7	45.1	68.3	65.2	94.5
Average Operating Cost	-0.1	-3.5	-25.9	-11.1	-33.1	-22.0	-25.8
Accident Rate	-41.2	-68.2	-76.0	-43.0	-60.8	-60.8	-46.0

Source: Mizutani and Nakamura (2001).

Conclusion

It is not clear how much of the Japanese success is due to the change of ownership plus a better design of subsidies, or to the market organization chosen in term of vertical and horizontal industry structuring.
The distinctive feature of this case is the choice of industrial organization. In the trade-off of vertical structure the loss of economies of coordination mattered more than the forgone gains of on-track competition. However competition is already present through the existence of private rail operators. Therefore, the high density of the network means the track is no longer considered an "essential facility", and it is economically feasible to have competition among vertically integrated structures. Horizontal separation is an indirect way to channel competition via relative performance regulation (yardstick competition).

Finally, it is not clear which elements of this experience are suitable for translation to other countries, whose conditions of demand and geography are not the same as in Japan. European countries have in general lower traffic density, as is shown in Table A3 below. However competition from other modes cannot be ignored.

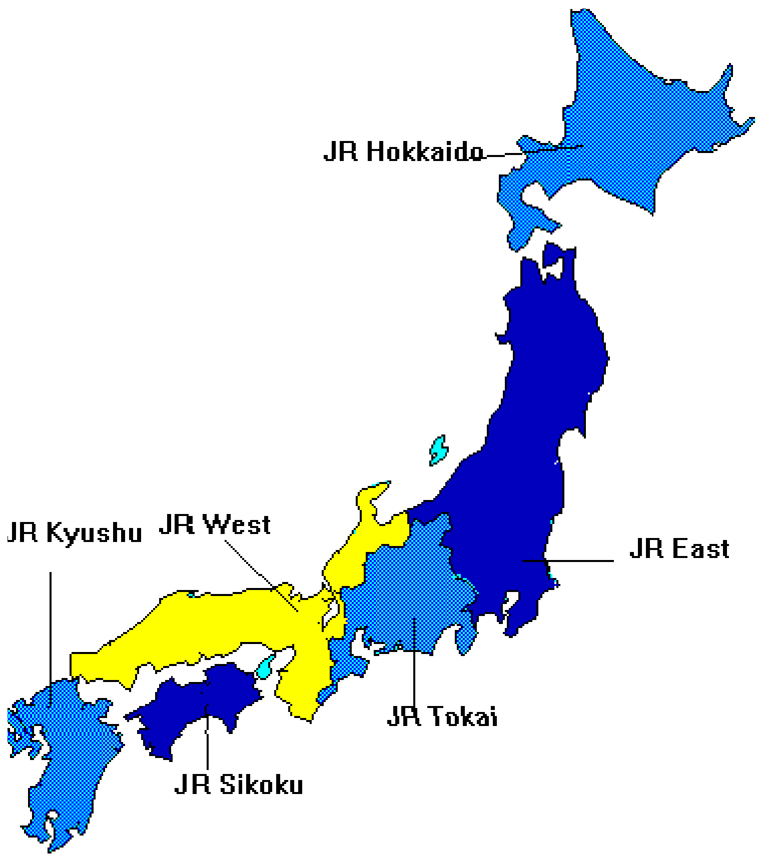


Figure A1: Areas Served by the six regional companies

Table A3: Traffic density for European and Japanese carriers (1995)

Firm	Density Pass-train km/km per year
BR (UK)	24.855
SJ (Sweden)	10.258
SNCF (France)	14.108
NS (Holland)	34.610
JR (Japan)	38.801

JR corresponds to the six regional networks.

APPENDIX 3: THE NEW ZEALAND CASE

The reform of the railroad industry in New Zealand was motivated by the poor commercial performance of the state-owned company.

This reform was part of the global change in the freight transport industry in the country. The freight transport was deregulated in the mid-1980 in terms of eliminating price–fixing and exclusivity of operation for railroads in some routes. The rail company was then privatized in 1993 as a vertically integrated operator. The objective pursued was to make this integrated rail firm compete with road and ship transportation in a deregulated environment.

There exist rights of access to the rail network but in a very restrictive way, in the sense that the access does not allow any interference with the normal operation of the rail company. There was no intention from the government to create or encourage on-track competition.

The outcome of the reform appears strongly positive under most reasonable measures. The company has returned to profitability, there has been an increase in productivity and rail market share and also users have been benefited due to the decrease in freight tariffs.

A Cost-Benefit study has estimated the welfare gains of the reform between 5.4 to 9.8 Billion USD. Taxpayers have been the main winner because they have not longer to provide public funds since 1993. Rail operators have not gained extra-normal profits compared with a diversified portfolio.

The case is of less relevance to DB than the others because it was confined to freight transport. Nevertheless, it indicates that vertical integration may be compatible with an efficient performance by a privatized company even in the absence of on-track competition, provided intra-model competition is strong enough.

APPENDIX 4: THE AUSTRALIAN CASE

The restructuring of Australian railroads is an on going process that started in 1990. Since then, several reforms in firms' governance, industry structure and regulation have been accomplished.

Characteristics of the Network

The Australian network is 43100 km long. The intensity of use is not comparable with those of European countries, which have a much higher population and network density as well (France rail network is around 35.000 km long). These facts make the use of a private car the privileged mode of transport even for urban trips.

The rail system can be divided into:

- The Urban Network is extended around big cities in the South. This network faces strong competition from other modes, especially private cars. Rail urban transport counts for around 4.0% of the market.
- The Non-urban Network includes Regional and Interstate lines. This system links the big cities with the inland areas. It is mostly used for freight. The market share of passenger travel has declined from 10% in 1970 to less than 2% in 1995. This transport market is dominated mostly by car travel (71%) and air (19%).

First Steps in the Reform

At the beginning of nineties, the railroads were state-owned and vertical and horizontally integrated at the level of geographical-political jurisdictions (States). The national government owned and operated the interstate line in passenger and freight and also some freight services within some jurisdictions.

The reforms in terms of industry configuration and operation were dissimilar across the states. Figure A2 presents a summary of the new structure of Australian railroads in 1999.

The more interesting cases are:

In New South Wales (NSW) services were horizontal and vertically separated. Four entities were created:

- Rail Access Corporation. Owns the track of the inter and intrastate network.
- Freight Corp. For freight train operations.
- Rail Services Authority. Provides Maintenance services.
- SRA. Provides urban and inter-urban passenger services.

In Victoria the service was horizontally separated. The freight business was separated from passenger service and privatized in 1999. The passenger transport remains vertically integrated, but it was separated in one non-urban and two urban services. These contracts were allocated to private operators through competitive tenders. The franchise terms include length of the license, investment commitment, maximum fare, and subsidy due by the authority among other things.

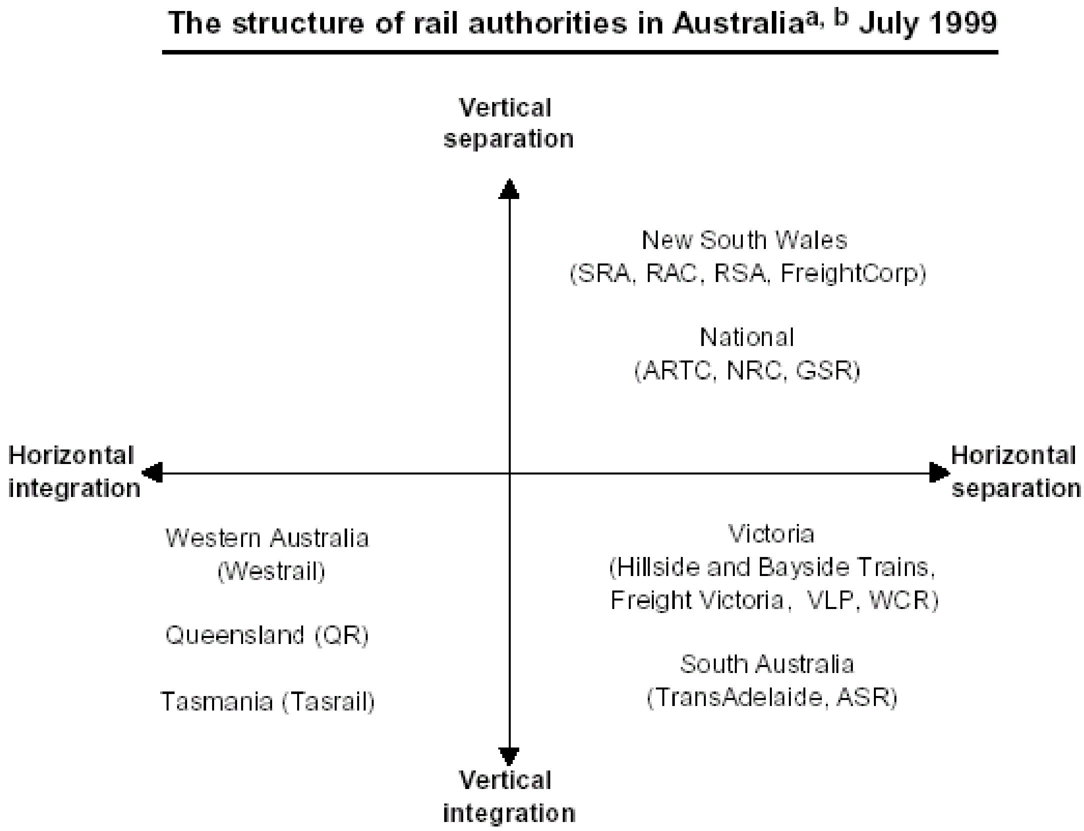


Figure A2: Australian railroad reform

Table A4: Urban and non-urban passenger franchises in Victoria

Franchise	Contract Length Years	Subsidy in 2000-01 M$	Investment M$
Bay side Trains	15	83	640
Hillside Trains	15	91	490
V/Line Passenger	10	78	165

Access to the Infrastructure

There is no uniform access policy. The flexibility in decisions reflects a wish to grant access only when this is likely to increase welfare, which is not always the case.

Firms seeking entry may either directly negotiate with the track owner (or administrator), or use the access regime established for the network or ask that the service be "declared". This last regime means that the National Competition Authority (NCC) recommends to the Ministry whether access to the network should be granted; the parties must then negotiate the terms of the entry under legally binding arbitration. Most of the States and the national network have defined their access regime. In NSW state, for instance, there exist price thresholds (ceiling and floor); other states establish conditions on prices, scheduling allocation and other variables. At the date of this report four companies (mostly freight operators) have applied for declaration. Access has been recommended in three of them.

Competition

So far, the reforms realized have employed mostly the concept of competition for the market as a way to provide competitive service in the passenger transport market (franchises in Victoria). The strong inter-modal competition that rail faces leaves no room for more than one feasible firm running the same service.

The current access regime leaves an open door for providing entry and some degree of desirable competition in the case of services having some level of market power.

Further Reforms Proposed

The structural reform proposed by the PC (Productivity Commission) is not a single proposal but a set of different solutions that are suitable for the different types of existing networks.

Urban Network

- Horizontal Separation from other business such as long-distance and freight;
- Vertical Integration infrastructure-operations;
- Allocation of the subsidized service by competitive tendering.

Interstate Network [22]

- Vertical Separation of infrastructure and train operations;
- Horizontal Integration. The entire interstate infrastructure should be managed by a single entity.

Passenger Regional Network

- Horizontal Separation based on geographical criteria.
- Vertical integration.

The Productivity Commission's Appraisal of the Australian experience

The Productivity Commission of Australia[23] has assessed the three main types of network (urban, interstate and regional) along the following dimensions:

- Interface issues, which occur when there are competing demands for train schedules by trains from different networks, for example, freight trains traversing urban passenger networks;
- Rail competition:
    - Competition 'for' the market: competition between bidders tendering to provide a given service; or
    - Competition 'in' the market: competition between train operators for the same customers.
- Different forms of competition:
    - intermodal competition, particularly from road and shipping
    - Competition at different levels of viability:
        - loss making — requiring continual government funding to either the track or train operations; or
        - earning a reasonable rate of return that can support future investment and maintenance; or
        - achieving sustainable monopoly profits.

[22] Nation-wide network connecting main areas.
[23] The Productivity Commission Report (1999).

The characteristics of the different networks are summarised in the following table and discussed briefly below.

Table A5: Characteristics of Australian rail networks

Network	Interface issues	Rail Competition		Intermodal competition	Level of viability
		For the market	Between train operators		
Urban	Yes, especially Sydney	Some	No	Yes	No
Interstate	Yes	Limited	Yes	Yes	Uncertain
Regional	Yes	Limited	Limited	Most freight	Uncertain
Main coal lines	Yes	No	No	No	Yes

Urban passenger networks

Interface issues

The potential for interface issues to arise varies considerably across states. Interface issues are of particular concern in Sydney where there is congestion on the urban passenger network restricting the passage of freight trains. This is due to the complexity of the network and the intensity of use by passenger trains at peak periods.

Rail competition

With the exception of Melbourne, there is no rail competition on urban passenger networks. In Melbourne, there is competition for the market in providing urban passenger services. National Express and Melbourne Transport Enterprise have secured the franchises to operate Melbourne's Bayside and Hillside Trains respectively. However, the successful franchisees do not compete over the same tracks for passengers. Instead, the franchises are based on geographic service groups.

Intermodal competition

Urban passenger networks are subject to strong intermodal competition from the private car and other transport modes. The private car has been cited as the greatest threat to public transport, undertaking around 95 per cent of all urban trips[24].

Other public transport modes also provide strong intermodal competition. The majority of bus, tram and ferry services move people from the suburbs to the central business district, sometimes providing commuters with more than one public transport option. As noted by the Productivity Commission:

"Aggregate figures [on private car and public transport mode shares for urban transport] do, however, conceal the importance of public transport for some types of journeys. For example, 52 per cent of commuter trips to Melbourne's central area are by public transport and 80 per cent of workers in Sydney's central city use public transport to get to work".

Level of viability

Urban passenger networks in Australia are loss making, requiring continual government funding. In New South Wales, the Government allocated around $1 billion in recurrent and capital

[24] Cox, J. (1997), Roads in the Community — Part I, Are They Doing Their Job?, published by AusRoads.

funding for SRA in 1997-98. In Queensland, the average government payment (subsidy) per urban rail passenger journey is over five times the average fare paid by passengers.

The interstate network

For the purposes of this inquiry, the Commission has defined the interstate network as that presented by National Rail Corporation: the standard gauge track linking all mainland State capital cities; the lines linking Sydney, Broken Hill and Crystal Brook; and the branches to Whyalla, Western Port, Port Kembla and Alice Springs.

Interface issues

As described earlier, with the exception of Sydney, interstate trains (freight and passenger) generally have limited interface with urban passenger networks. This situation is primarily due to the differences in track gauges. However, as described later, there are interfaces between the interstate and regional networks.

Rail competition

There is competition between train operators on the interstate network. In June 1995, SCT, commenced interstate rail freight operations in competition with NRC. TNT (now Toll Rail) followed one year later. There is only limited evidence of competition for the market on the interstate network. Rail Access Corporation (NRC) noted that competition for the market has occurred on the interstate network.

Intermodal competition

There is vigorous intermodal competition on the interstate network, especially from road transport. As argued by NRC, road transport is the most powerful competitor for rail general freight services. Apart from competition from road transport, coastal shipping also dominates the interstate transport of bulk commodities.

Level of viability

The Commission received no evidence of railways extracting monopoly profits from customers on the interstate network. Indeed, there is no conclusive evidence of the ability for railways to achieve viability, at least at this stage. In terms of train operations, NRC continues to run at a loss. The Australian Rail Track Corporation illustrated the challenges facing rail by reference to NRC:

"NR's financial profitability was stated at $4.8m (loss) in 1996-97. The most recent annual report, released late last year, shows that NR's operating loss has deteriorated to $9m (after-tax) … This deterioration continues a trend starting in 1995-96 (with the introduction of private rail competitors) and starkly illustrates the challenge NR faces in a competitive interstate environment."

However, NRC argued:

"As a corporation under the Corporations Law, National Rail can trade only while it remains solvent. Since it ceased to receive any financial support from its shareholders some 15 months ago, its Directors must have an expectation of commercial returns, backed by shareholder-approved strategic plans."

As noted earlier, there are now private sector operators on the interstate network. As private firms, SCT and Toll Rail would be expected to only remain in the market if they earn, or expect to earn, commercial returns (at least on train operations). Evidence on the viability of the interstate track is limited due to the multiple owners of the network. However, in New South Wales the Government provides subsidy payments to RAC towards track upkeep and maintenance.

Regional networks

The Draft Report identified two types of regional networks — high and low volume. This categorization tended to create some confusion over what tracks could be considered high or low volume. This was especially evident in New South Wales where some tracks were used by trains carrying freight with different economic characteristics (coal, grain and containers), as well as non-urban passenger trains.

Nevertheless, most of the tracks within regional networks[25] share similar economic characteristics. The freight transported is often subject to strong intermodal competition and there is no evidence of railways extracting monopoly profits.

Only certain lines, namely the main coal lines in New South Wales and Queensland, display distinctly different characteristics to the rest of the regional network.

Tonnages of coal carried over the Hunter Valley coal lines exceed 50 million tonnes (Mt) per year (PC 1998a). In Queensland, the two main coal lines are those centred on the Oaky Creek and North Goonyella regions (around 49 Mt per year) and Gregory and South Blackwater regions (around 24 Mt per year).

The transport of coal and minerals in Australia (excluding the Pilbara) is not restricted to those lines identified above. However, the main coal lines are distinguished from the remainder of the regional network by the tonnages of freight that are transported over the lines and, as discussed later in this section, the ability of railways to extract monopoly rents from mining companies.

Trains transporting grains, general freight and non-urban passengers also travel on the main coal lines. Despite this overlap, the main coal lines have different economic characteristics and specific issues not associated with the remainder of the regional networks in New South Wales and Queensland. Policies required improving the outcomes for users (primarily mining companies) on the main coal lines are therefore identified separately.

Interface issues

Trains commencing on a regional network carrying minerals, grain or general freight often traverse both the interstate and urban networks. There are also interfaces between the main coal lines and the remainder of the regional network. As noted by the NSW Minerals Council:

"In hauling coal from Gunnedah or Ulan to Newcastle, the coal is first hauled on a low volume regional network, then on a high volume regional network, then on track that is used by and influenced by interstate freight and urban passenger traffic."

Rail competition

With the exception of New South Wales, there is limited evidence of rail competition either for the market or between train operators on regional networks. An example of competition for the market in New South Wales is the contract won by NRC to carry coal to Macquarie Generation's Bayswater and Liddell power stations. Another example of competition for the market (outside New South Wales) was the competitively tendered contract to haul coal from Leigh Creek to Port Augusta won by FreightCorp in November 1998.

A number of participants argued that rail competition between operators was beginning to emerge under the vertical separation model adopted in New South Wales. RAC argued:

"… approximately two-thirds of freight in New South Wales on a gross tonne-kilometre basis has been subject to competition between operators …"

A similar view was shared by NRC:

"It is not correct that "There is little or no competition either for the market or between train operators on low volume regional railways". Instances of competition on low-volume lines are very

[25] Regional networks are defined as those lines extending out from capital cities and regional ports to inland areas, excluding the defined interstate network and private railways in the Pilbara.

few as the possibility of competition on regional lines is very recent; the NSW Rail Access Corporation has provided successfully for genuine competition and in NSW there are several examples where real competition has occurred."

However, any competition between train operators in New South Wales will largely occur on subsidized track.

Intermodal competition

In Victoria, Western Australia (excluding the Pilbara region), South Australia and Tasmania, the majority of freight carried by rail on most regional networks is subject to strong intermodal competition. In terms of grain transport in Victoria, Vicgrain noted that it has:

"... the option of increasing the road component of its grain movement operations. Vicgrain envisages that this would occur should freight rates increase or should there not be a suitable level of cooperation between the storage and rail sectors."

Westrail argued that the majority of the commodities it transported by rail was subject to competition from road transport, with the exception of bauxite, where road transport was a less feasible alternative. Both FreightCorp and Queensland Rail have indicated that they face significant competition from heavy road vehicles in the general freight market. However, RAC argued that the transport of a number of commodities by rail on the regional network in New South Wales was not subject to intermodal competition:

"There are some traffics on low volume regional lines, particularly minerals and to a lesser extent grain, that are not subject to significant intermodal competition."

Similarly, the main coal lines in New South Wales and Queensland face little or no intermodal competition in the transport of coal.

Level of viability

Excluding the main coal lines, the Commission received no evidence of railways extracting monopoly profits from customers on regional networks.

In New South Wales and Queensland, the regional networks can be considered loss making, requiring subsidy payments in excess of $150 million per year. Despite the lack of intermodal competition highlighted by RAC, there is no evidence of railways extracting monopoly profits from customers in New South Wales. An important fact underpinning this position is the high level of competition in the final markets of these commodities. Commodities such as export grain face strong competition from alternative suppliers on international markets. Thus there are no monopoly profits to be earned by grain farmers in final markets to be extracted by the providers of inputs, including transport.

This conclusion is reinforced by the fact that the rail transport of commodities such as grain in New South Wales receives government subsidies. If grain farmers require government subsidies to cover transport costs, it is unlikely that monopoly profits are simultaneously being extracted by the railways.

In Victoria, Western Australia, South Australia and Tasmania there is some evidence to suggest the regional networks can earn a reasonable rate of return. With the exception of Western Australia, new (vertically integrated) private sector railways have begun operation (Freight Victoria, Australia Southern Railroad and Tasrail). These private sector companies operate without government subsidies and would only enter and remain in the market if they earn, or expect to earn, a commercial return.[26] With regards to the viability of rail freight operations in Tasmania, the Australian Transport Network (owners of Tasrail) stated:

"... since acquiring Tasrail, its revenue has increased by approximately 50 per cent and the company has been returned to profitability."

In Western Australia, Westrail receives no subsidies for freight operations. However, Westrail does require subsidy payments for non-urban rail passenger services (some $13.7 million 1997-98).

The main coal lines in New South Wales and Queensland are the most profitable components of each State's network. The extraction of monopoly rents from the transport of coal by the government-owned railways provides indirect evidence of the profitability of these lines.

Overall appraisal of Australia's rail networks

Australia's railways provide transport services to a diverse range of passenger and freight markets. However, the discussion above has highlighted some distinct characteristics that exist between States with regard to urban and regional networks. No urban passenger service in Australia is viable. Participants highlighted the problem of congestion between freight and passenger trains on the urban passenger network in Sydney due to its complexity and the intensity of use by passenger trains at peak periods. On the other hand, in Western Australia and South Australia, the urban passenger networks have limited interface with other networks.

Notwithstanding some possible exceptions, regional networks (excluding main coal lines) are characterised by strong intermodal competition, especially from road transport. Where intermodal competition is absent, there is usually competitive pressure in downstream markets limiting the ability of railways to extract monopoly profits.

The regional networks in New South Wales and Queensland are distinguished from the other States by the considerable government subsidies required to ensure their continued viability. While some competition between train operators is emerging in New South Wales, these operators do not cover the full cost of providing track infrastructure.

Yet the New South Wales and Queensland regional networks also contain the identified main coal lines that are distinguished by their profitability.

[26] In Victoria the State Government provides some limited subsidy payments for the 'Fast Track' service (less than container load freight).

APPENDIX 5: THE UNITED KINGDOM CASE

Railways have been an important network industry since their emergence in Britain early in the 19th Century. More recently, Britain can claim to have pioneered organisational reorganisation of its rail industry, as a result of the privatization that took place in the 1990s.

Initial Conditions

Railway services in Britain were nationalised in 1948 following the 1947 Transport Act, which created the British Transport Commission. The 1962 Transport Act saw the creation of a separate British Railways Board, which remained in public ownership. Criticisms of the British Railways Board began to emerge in the 1970s, highlighting low productivity, inappropriate investment, managerial inefficiency and spiralling subsidies.

The commercialisation of BRB in the 1980s led to some marked improvements in productivity - however, these improvements took place against a backdrop of privatisations and deregulation in the rest of the economy. Thus, calls for the privatisation of British Rail as well were inevitable. The search for a privatisation formula gained momentum in the early 1990s as BR's productivity and financial performance began to deteriorate again and the Conservative privatisation programme started to run out of steam.

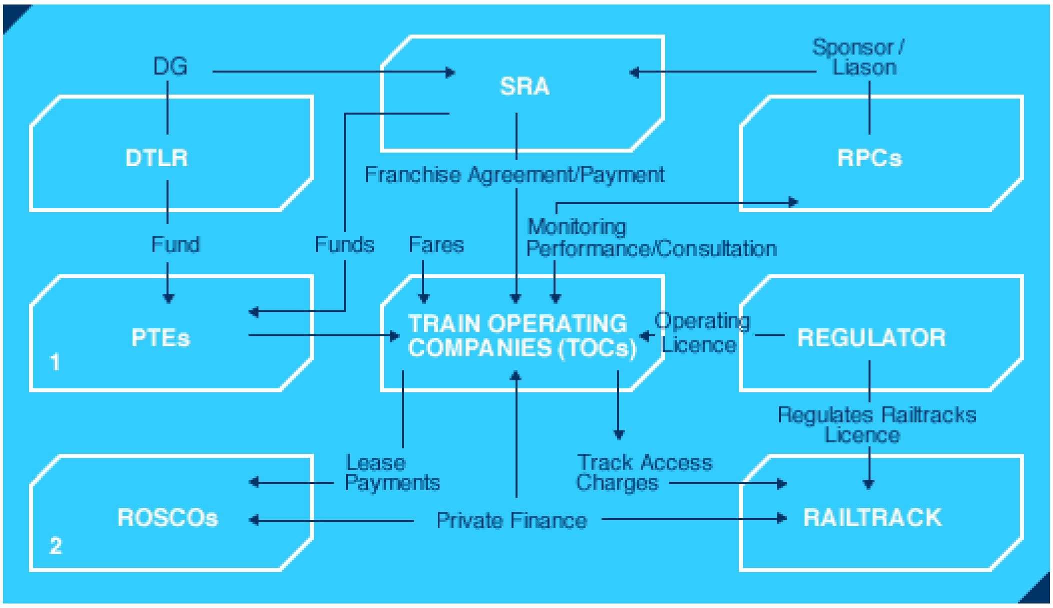
Structure

In 1992, the Conservative government released a plan for privatisation of British Rail, in the form of a White Paper entitled 'New Opportunities for the Railways. The 1993 Railways Act and subsequent Government directives largely implemented this plan.

Following Labour's victory in the 1997 election, the industry was partially re-organised. Specifically, the Strategic Rail Authority was set up, changing the regulatory structure of the industry.

The UK Rail network was vertically and horizontally disintegrated, in a process which separated infrastructure from operations. The industry now consists of four main operational components: originally 25 Train Operating Companies (TOCs); 5 freight operators; 3 Rolling Stock Leasing Companies (ROSCOs); and the Rail infrastructure operator, originally called Railtrack, which was privatised in 1996 but has now been taken back into public ownership and is called Network Rail. Regulation of the industry is achieved by a curious two-regulator model.

The present structure of the UK Rail industry can be represented diagramatically:



Source: SRA website.

Figure A3: Structure of UK Rail Industry

Rail infrastructure was separated from rail operations and is now the responsibility of Railtrack, initially a separate government-owned company but subsequently floated on the stock market in April 1996. In 2001, following a train crash at Hatfield, it was discovered that the state of the rail infrastructure was considerably worse than had previously been thought. Railtrack set about a massive project of rail replacement, which was both extremely expensive and very disruptive to the operation of the rail network. The cost of financing its increased level of debt became too much for Railtrack in February 2002, and the Transport Secretary Stephen Buyers took the company into receivership.

The passenger rail business of BR was split into 25 train operating units, broadly corresponding to existing profit centres. Following privatisation, the 25 passenger franchises were sold to 11 different parent companies. The duration of the first franchises was set at 7 years, so re-franchising negotiations are currently in progress. So far only one franchise has been re-negotiated, with a new 20-year franchise being awarded to Chiltern Trains.

Several freight operators were privatised: Trainload Freight, which specialises in the carriage of bulk raw materials, was divided into three geographically based companies; Railfreight Distribution, was privatised in two parts – Freightliner, its domestic container business, which transports containers to and from UK mainland ports, and the European intermodal and automotive freight business; Red Star, the express parcels service, was sold in 1995 as a MBO; Rail Express Systems Ltd., whose principal business is the carriage of mail for the Post Office was sold in 1995.

Three rolling stock leasing companies (ROSCOs) have been established: Angel Trains, Eversholt Leasing and Porterbrook Leasing. Each ROSCO owns between 3000 and 4500 vehicles, consisting of a mixed portfolio of different types and ages of rolling stock. The ROSCOs own and lease all of the domestic passenger rolling stock previously owned or leased directly by British Rail.

The regulatory structure of the industry is as follows. The Office of the Rail Regulator is in charge of granting licences, which are required for the operation of certain railway assets. There are four main types of license: train operators' licence (passenger or freight licence); station licence; depot licence and network licence. The basic position is that Railtrack awards safety certificates to train operators and the Rail Regulator certifies Railtrack.

The second tier of regulation was initially provided by the Office of Passenger Rail Franchising (OPRAF). This was a non-ministerial government department, which awarded rail franchises to TOCs. Its role was: to award franchises on the basis of competitive tendering; to provide for the delivery of specified assets at the end of the franchise; to minimise disruption of services; to ensure that fares charged by a franchise operator are reasonable and that operators participate in approved discount schemes for the young, elderly and disabled; to receive and give notice of proposals to discontinue passenger services; to approve the designation of certain services as experimental; to secure compliance with the terms of franchise agreements.

The Strategic Rail Authority, set up by the incoming labour government, replaced the OPRAF, but took over most of its duties. Its purpose is to provide a focus for the planning of passenger and freight railways. It supports integrated transport initiatives and for the first time provides a clear focus for the promotion of rail freight. The SRA has unspecified funds to support its activities – its potential sources of revenue include Railtrack, the TOCs, and the Treasury's capital and revenue support.

The SRA's main role is consumer protection at the franchise level, with its secondary objectives being to: promote the use of rail within an integrated system; ensure it is planned as a coherent network; participate in regional and local land-use planning policies; ensure that rail options are assessed to achieve value for money and optimise social and environmental goals; take a view on the capacity of railways, assess investment needs and identify priorities; promote accessible transport for the disabled; review the role for rail in sustainable transport; produce criteria for competition between operators.

Following the creation of the SRA, the Office of the Rail Regulator's main role became the regulation of Railtrack, the ROSCOs and the implementation of competition policy.

The original privatisation plan envisaged two types of access agreements for passenger services: franchised passenger services access agreements; and open access agreements where operators would be able to negotiate open access agreements with Railtrack. It soon became clear that encouraging on-track competition would conflict with the government's objective to reduce subsidies.

Thus the Rail Regulator has devised a regime of 'moderation of competition' for the passenger franchises. This regime determines in advance the degree of exposure to competition which each franchisee could face, ensuring that the adverse impact on the franchising process would be minimised. New entry is restricted through contractual control over Railtrack's ability to sell access rights to any operator in addition to those contained in their initial access agreements. The mechanism was intended to allow the market to determine as far as possible where competitive entry should occur. The protection given was to be over the markets served by the operators rather than the track on which they operate.

The mechanism operates in two stages. The first started when the last British Rail TOC has its long-term track access agreement approved and was originally planned to expire on 31st March 1999. The second was planned to expire on 31st March 2002.

- First stage of open access

In the first stage, franchise holders can nominate a list of substantial point-to-point flows on which no new entry for scheduled passenger services will be permitted without the franchisee's agreement. Franchise operators can nominate additional flows, not conforming to this rule, subject to the approval of the regulator. This is designed to allow passenger operators additional rights where there is no conflict with the protection given to other operators.

- Second stage of open access

In the second stage, some new competition via open access is intended to be introduced. As with Stage 1, the key building block is the station to station flow and franchisees will be expected to re-nominate flows, but in Stage 2, franchisees will only be allowed a partial protection from new entrants. Up to a threshold level set by the Regulator, new entry can then occur, and compete for the operator's business.

In stage 2, the competition mechanism is designed to limit the amount of the operator's passenger revenue that is open to competition from another operator.

Conduct

The everyday interactions between Railtrack, the ROSCOs and the TOCs are extensively controlled by the regulatory regime. Thus, the conduct of the industry is inseparable from the question of regulation. Thus the regulation of Railtrack, the ROSCOs and the TOCs will be discussed in the following section. The regulatory framework set up during privatisation has been much criticised, essentially because the underlying incentives put in place were inappropriate.

The conduct of the UK rail industry takes place in very many dimensions. Railtrack/Network Rail takes short run decisions on the amount of money to spend on track maintenance and who perform this maintenance. It also takes long run decisions on investment in replacement tracks, signalling and stations, as well as the provision of new routes. The TOCs choose the number of trains to run, the times to run these trains, the type of rolling stock to use, the prices to charge. They also plan new services, in conjunction with both Network Rail and the ROSCOs.

Network Rail

Network Rail interacts mainly with the TOCs, which operate rail services using infrastructure owned by Railtrack. Both the conditions under which Network Rail is obliged to grant access to its infrastructure, and the access prices which it can charge are controlled by the Office of the Rail Regulator. The most important segment of Network Rail's earnings are passenger services: around 90% of its income is from passenger operators. Thus the Regulator can have a very significant impact upon the financial performance of a company. Network Rail has more flexibility in dealing with freight traffic because it works with freight operators to schedule trains outside peak periods.

Network Rail operates a track-charging system which was put in place during privatisation, in conjunction with the Rail Regulator. Each rail user has to pay Network Rail based on the following formula:

(i) A fixed annual charge.
(ii) A variable track usage charge based on the way maintenance costs vary with vehicle miles, speed and axle weight. Overall this component accounts for 3% of costs.
(iii) A charge for electricity usage, based on vehicle miles, train miles, or gross tonne miles. Overall this component accounts for 6% of costs.
(iv) An adjustment to reflect additional costs incurred and savings made by Network Rail as a result of a change of law, for example new environmental legislation.
(v) An adjustment to reflect the financial consequences for Network Rail of managing or dealing in property.

Regulation of Railtrack was originally set up upon the traditional utility regulatory model – the ORR would regulate Railtrack through a conventional RPI – X regime, supported by a licence. In the initial period following privatisation, the ORR was obliged to discipline Railtrack due to its poor performance. However it was somewhat constrained in doing this by the sketchiness of its initial licensing agreements with Railtrack. In response to criticism and because volumes turned out higher than expected, the early notions of Railtrack managing a static railway network were replaced by a more demanding set of infrastructure renewals and enhancements. The absence of information on asset condition, and the lack of much definition in the contracts for maintenance between Railtrack and its main contractors, made the job of both the company and the Regulator very difficult in the first period.

Because a large proportion of the Train Access Charges are fixed, Railtrack had an almost guaranteed income from the TACs. Put simply, Railtrack's profits were the difference between that income and costs. Volume was not of much interest – indeed, more passengers and freight could actually make matters worse by increasing costs. Penalties for non-delivery were trivial in comparison to the difficulties of meeting higher levels of demand.

Thus, the incentive regime put in place led Railtrack to concentrate on short run conduct variables, essentially on squeezing costs. Thus Railtrack reduced its workforce by contracting out maintenance services.

Few long-term investment schemes have been embarked upon, because the incentives for such schemes are weak relative to the risks of failure. The few big investment projects which have gone ahead have been dogged by delays and cost over-runs. The West Coast Main Line improvement is a case in point, and will be discussed further in the 'Performance' section. The essential point is that while a price-cap regime has worked relatively well for mature networks such as electricity and gas, the rail industry is an investment-driven business where tough price regulation may be incompatible with attracting funds for investment. Therefore it is unclear that it is appropriate to place tough operating cost reduction targets on a company which should be more concerned with performance and investment than cutting costs and jobs.

The creation of the SRA was intended as a partial remedy for the lack of long-term investment on Railtrack's part. The SRA has a mandate to consider the long-term development of the rail industry. However, given that the SRA does not regulate Railtrack, it is clear that the institutional setup is sub-optimal to achieve these objectives.

Less importantly, Network Railtrack also charges for access to stations and depots. Most stations are leased to individual train operators, although some large stations are run separately by Network Rail.

Train Operating Companies (TOCs)

The Conservative model of privatisation envisaged that TOCs would only need temporary regulation. Access to the rail infrastructure would gradually be opened up, so that TOCs would enter into competition on passenger routes. Competition would ensure that TOCs could be left to choose their own conduct variables without preying on consumers.

This proved to be unrealistic. Open access to the rail infrastructure remains a distant prospect, so most passenger routes are operated by monopolies, thus necessitating regulation. But the privatisation plan did not anticipate this need for regulation. As a result, regulation of the TOCs was not built into the ORR's mandate, and instead was deal with by the OPRAF, and then later by the SRA. Therefore the regulation of the Rail industry is divided between two entirely separate regulators.

While there is very little on-track competition, there does exist off-track competition, through the franchising process. The franchising process assigns monopoly-conferring property rights to competing bidders. The terms of the franchise are set subject to several requirements, along the dimensions of price and output. The current framework includes: a passenger service requirement, stipulating minimum levels of service; a price cap; operating performance measures; payments to ensure that TOCs have effective incentives to improve performance.

(i) Passenger Service Requirements (PSR)

The Passenger Service Requirement is to ensure the provision of a minimum service level, while allowing the franchise operator to adjust its timetable to passenger requirements, thereby improving efficiency. Thus the PSR sets parameters within which each TOC must design a timetable.

The PSR specifies service characteristics which are important for passengers including: frequency of trains; stations to be served; maximum journey times; first and last trains; week-end services; through services; and load factors and/or peak train capacity.

(ii) Fares

Constraints are placed on:

- Unrestricted standard class return fares and certain single fares for short distance journeys;
- 'Saver' fares on other journeys;
- certain standard class season ticket fares including all those tickets for weekly season tickets.

For three years from 1st January 1996, increases in capped fares were not permitted to rise by more than the Retail Price Index from the 1995 base price. For the four years from 1st January 1999, the price cap for fares has been RPI - 1. There is some limited scope for individual fares to exceed the cap where these are balanced by other controlled fares being held below cap levels. The caps on the price of fares in the London area are adjusted to reflect the quality of the franchise operator's performance, in order to ensure that passengers are both compensated for poor performance and are made to pay for improved performance through fares.

(iii) Operating performance

Each franchise operator was required to produce its own Passenger Charter, which must be at least as good as that formerly in place for British Rail. In addition, all franchise operators are required to conduct regular, independent customer satisfaction surveys. Bad performance results in some repayments to the passengers. Breaches of obligations result in the Franchising Director setting out the action needed for compliance. Exceptionally poor performance can lead to a special kind of breach, an event of default. This gives the Franchising Director the option to terminate the franchise.

The Passenger Charter has now been replaced by the "Public Performance Measure", namely the percentage of trains arriving on time. As will be seen below, this measure has deteriorated substantially over time.

iv) Franchise payments

The franchise payments paid by the Franchising Director to the TOC aim to put in place an incentive regime whereby operators have a financial incentive to improve performance.

- The Punctuality Incentive Payment (PIP) measures lateness and cancellations against the planned timetable. The results for each four week accounting period are compared with the benchmark figure established before franchising.
- The Short Formation Incentive Payment (SFIP) applies where the Passenger Service Requirement specifies a minimum capacity that the operator has to provide. This applies for peak services into London and some commuter flows to other major cities. Each operator has to have a train plan showing how the capacity will be delivered. If the operator fails to meet the train plan, then an SFIP charge, based on a proportion of the cancellation charge, is made.
- The Timetable Change Incentive Payment (TCIP) penalises operators who change the timetable from the printed version. Because it substitutes for a higher payment under PIP,

the TCIP gives operators an incentive to handle disruption in a planned way, and to give passengers advance notice of amended services.

It is clear that the regulatory regime surrounding the franchising process places very many constraints on the TOCs. Indeed, one could say that most of their short-term conduct involves trying to cut costs while remaining inside the performance guidelines set out above. Cutting costs generally involves running the minimum number of trains allowed, while controlling labour costs by employing few workers at low wages. They also have some scope to minimise the costs of leasing rolling stock through a competitive tendering process with the ROSCOs.

Longer-term conduct variables include the introduction of new services by the TOCs. The TOCs have consistently argued that the franchise periods were too short to encourage them to invest in new trains and to upgrade services. Uncertainty over whether they would retain their franchise in the future led TOCs to concentrate on short-term goals. However, as always with franchising, the relevant distinction is between capital costs that are fixed and those which are sunk. Sunk costs cannot be recovered on exit. While this may have provided a disincentive to invest in improvements to stations (which remain attached to the franchise), TOCs in fact have very few sunk costs. Thus any distortion of time horizons away from long-term investment is probably more a fault of the incentive structure embodied in the franchise regime – i.e. are TOCs allowed to make appropriate revenues to recoup the costs of setting up new services.

Freight Operators

The rights of access to the railway were made available to private freight operators without a franchise, but restricted for passenger services up to 2002.

ROSCOs

The ROSCOs were sold by the government in 1996 for around £1.82 billion, and in a period of less than two years then had been sold on for around £2.65 billion.

When the TOCs want to lease rolling stock, the ROSCOs compete for the contract. As for new rolling stock, the TOCs renew their rolling stock in consultation with the manufacturers and the ROSCOs. Contracts between ROSCOs and TOCs are generally of up to 10 years duration.

Performance

Gross Output

Gross output levels give an indication of the raw output of the passenger train industry. From a low of 349 million train kilometres in 1992/93, the train kilometres increased by 3.2% per annum until 1996. From 1997-2001, train kilometres increased by over 5% per annum. However, one should realise that passenger demand increased substantially in the between 1995-2001 due to rapid economic growth in that period, combined with a slowdown in the growth of private car ownership, a significant real increase in fuel costs for motorists and increasing congestion in populous parts of the country.

Quality of Outputs

(i) Punctuality

Some initial improvements in the punctuality of trains occurred after privatisation, but this improvement turned into a marked subsequent deterioration. 89.7% of trains arrived on time in the financial year 1997-8, but by 2001-2 this had fallen to 78.0%.

(ii) Reliability

Reliability improved consistently during the 1990s, although like punctuality, most TOCs have shown a decline in reliability for 2001, compared with the previous year. Although many of the problems concerning punctuality coincide with known infrastructure problems, such as poor track and signalling quality, lack of track capacity and lack of terminal capacity, it is reported that the percentage of delays attributable to Railtrack has declined from 70% in 1995 to 43% in 2000.

(iii) Congestion

Due to demand increases, overcrowding has also increased during the last five years. Last year, congestion was is in part caused by the disruption of speed restrictions and severe weather, and incentive penalties were paid by TOCs as a result. Ideally, the passenger regulator should calculate the optimal level of congestion for rail services, and then set overcrowding penalties accordingly. However, one suspects that the regulator may err on the side of too little crowding, so it is difficult to interpret recent increases in congestion.

As a reflection of the overall quality of outputs, the SRA received net incentive payments of £103.3 million from operators in the year ending 31 March 2001. This compares with a total net payment made by the SRA to the operators of £0.87 million in the previous year.

New Services

In the period following privatisation, around 113 key franchise commitments were proposed. By 2001, 45 of these had been delivered:

- Of 35 service commitments, 21 were delivered by 2001, mainly related to maintaining the current timetable for a set period, whilst new and upgraded services have largely to be delivered.
- 20 commitments to improving stations. These were normally posited as a rolling programme so only two had been completed by 2000.
- 14 raising of Charter Standards, of which 4 have delivered. Typically this involves increasing punctuality and reliability targets, but not necessarily performance.
- 11 commitments to improving links with buses and other modes, the majority of which have been delivered.
- 12 other miscellaneous commitments, 8 of which have been delivered.

New Trains

There have also been 22 commitments to new or refurbished trains. 7 of these have been delivered. However, several of the plans for introducing new trains also involve the installation of new rail capacity, which is the responsibility of Railtrack. The largest of these projects involved commitments made by Virgin to introduce high speed tilting trains on the West Coast main line. In 1998, following negotiations with Virgin, Railtrack committed itself to upgrade the West Coast main line to allow trains to run on the fast lines at 125 mph by May 2002 and 140 mph by May 2005. However, delays and cost over-runs multiplied, and Railtrack subsequently sought an increase in funding through access charges to cover the estimated increase in costs from £2.1 billion to £5.85 billion. In 2002 the future of the project looks more uncertain than ever.

Fares

Inflation-corrected fares have remained reasonably stable since privatisation. Approximately 39% of train operators' ticket revenue comes from fares regulated by the SRA, which since January 1999 have been limited to RPI - 1. For the ten London commuter operators, the cap on their fares basket is adjusted by up to 2% above, or up to 2% below, the RPI - 1, under the Fares Incentive Adjustment Payment. Worsening train service performance in 2000 further reduced the increase in London commuter fares in 2001 and in many cases fares were reduced in real terms. On average these ten operators were allowed to increase their fares by just 1.3% – a 2% reduction in real terms. However, unregulated fares have risen sharply in recent years, as table A6 indicates:

Table A6: Fare increases 1999-2002

Fare	Percent Change January 1999-January 2002
First class	23.6
Standard class unregulated	11.0
Standard class regulated	0.8
Average All Fares	8.0
Retail Price Index (for comparison)	6.1

Source: Strategic Rail Authority, National Rail Trends.

Employment

Staff numbers decreased sharply during the mid-1990s, until stabilisation in the late 1990s. However, thousands of employees have been switched to railway-related industries, rather than being fired from the railway industry altogether. The cuts made led to substantial cost reductions, although some operators reduced operating staff numbers too much, and faced driver shortages as a result.

As a result of employment cuts, there was a large increase in labour productivity since the reforms. However, these figures may be artificially inflated by a statistical, rather than a real decline in employment. Thus productivity increases could be substantially smaller.

Investments in Infrastructure

During the years before privatisation, investments in infrastructure were low. Railtrack embarked on an extensive modernisation programme, focussing on signalling systems and control centres. However, the past two years in the rail industry have seen unparalleled disruption as Railtrack imposed hundreds of emergency speed restrictions around the network as it investigated the extent of ‘gauge corner cracking’. It was this ‘gauge corner cracking’ which was responsible for a fatal high-speed derailment at Hatfield on October 17th, 2000. Assessment of the situation led to a recovery programme, which required the replacement of many hundreds of miles of rail.

The year 2000-2001 showed the beginnings of a fall in the overall numbers of broken rails. As at 31 March 2001, results indicated that the annual total would be significantly below the prevailing level of the last two years. However, the need to impose such a large number of speed restrictions across the network following Hatfield highlights wider questions about Railtrack’s network stewardship, the knowledge of its assets, and its renewal and maintenance policies.

Safety

Since privatisation, over fifty people have been killed in railway accidents, and several hundred injured. Some have suggested that the vertical disintegration of the rail industry, combined with a focus on cutting costs, has led to more fatal accidents. However, there were a similar number of accidents in the 1980s, when British Rail controlled the system.

Overall Conclusion

Although there has been an increase in entrepreneurial activity in the rail industry following the reforms, the main focus has been on cost-cutting. Greater emphasis may be placed on growing the market in the long run, with the development of new services and the ordering of new rolling stock. In other words in the future passenger rail companies may be expected to move from strategies based on cost leadership to strategies based on differentiation and focus.

APPENDIX 6: REGULATORY REGIMES AROUND THE WORLD

Table A7: Part one

	Ownership type	Financial rehabilitation	Infrastructure	Performance criteria (for public authorities)	Public Service Obligation in the national legal context	Access rights	Regional rail transport
Austria	Infrastructure: Public Operations: Public	Not realized.		Safety levels.	Puts much more in multiyear-contracts with the railway undertakings. Uses a framework contract and a yearly contract where the precise service levels and financial compensation is regulated.	Access rights for long periods: 25 years.	One or more regional railway undertaking other than the incumbent operator. Experience with tendering for PSO contracts Budget responsibility at regional level.
Belgium	Infrastructure: Public Operations: Public		Construction, maintenance and management of the rail infrastructure: NBMS. Infrastructure financing: state or NBMS. User charges: YES	PassKm, Punctuality, Expected financial results, Safety levels.	Puts much more in multiyear-contracts with the railway undertakings.	Unlimited exclusive access rights.	No railway undertakings especially for regional passenger transport.
Denmark	Infrastructure: Public Operations: Public			Number of passengers, Public opinion statements from surveys, Financial performance of the railway undertaking (financial results not available on PSO contracts), Safety levels.	Puts much more in multiyear-contracts with the railway undertakings.	Access rights linked to the public services contracts, 5 years.	Plans to tender out regional rail passenger public service obligations contracts. Experience with tendering for PSO contracts. Budget responsibility at regional level. Planning for tendering PSO contracts.
Finland	Infrastructure: Public Operations: Public			Number of pass, TrainKm, Expected financial results, Safety levels,	Puts much more in multiyear-contracts with the railway undertakings.	Unlimited exclusive access rights	One or more regional contract with regional public authorities carried out by the incumbent operator. Budget responsability at regional level.

Table A7: Part one (continued)

	Ownership type	Financial rehabilitation	Infrastructure	Performance criteria (for public authorities)	PSO in the national legal context	Access rights	Regional rail transport
France	Infrastructure: Public Operations: Public	The debt situation has been cleared.	Close relationship between RFF and the political body. Responsibility of political bodies for rail infrastructure planning and financing.	PassKm, Safety levels, Financial results (not available on PSO contracts). Explicitly mentioned that regional traffic was growing faster in areas where this traffic was actually decentralized.	Puts PSO in the law or secondary legislation and compensation is established year by year. No public service contract does exist.		One or more regional contract with regional public authorities carried out by the incumbent operator. Budget responsibility at regional level.
Germany	Infrastructure: Public Operations: Public	The DB-AG was freed from old debt.		Safety levels,	Puts PSO in the law or secondary legislation and compensation is established year by year. No public service contract does exist.	Unlimited non-exclusive access rights.	One or more regional railway undertaking other than the incumbent operator. Experience with tendering for PSO contracts. Budget responsibility at regional level.
Greece	Infrastructure: Public Operations: Public						
Italy	Infrastructure: Public Operations: Public		Vertically integrated within FS-Spa that in turn is owned by the Treasury. Allocative inefficiency and policy driven investment decisions.		Puts much more in multiyear-contracts with the railway undertakings.	Unlimited non-exclusive access rights.	Plans to tender out regional rail passenger public service obligations contracts. Budget responsibility at regional level.

Table A7: Part one (continued)

	Ownership type	Financial rehabilitation	Infrastructure	Performance criteria (for public authorities)	PSO in the national legal context	Access rights	Regional rail transport
Ireland	Infrastructure: Public Operations: Public			Management accounts, passenger volumes and some internal performance indicators.			
Luxembourg	Infrastructure: Public Operations: Public		The state has the political and financial responsibility for the rail infrastructure.	PassKm, Safety levels,	Puts much more in multiyear-contracts with the railway undertakings.		No railway undertakings especially for regional passenger transport.
The Netherlands	Infrastructure: Public Operations: Public + Private		The government is responsible for policy n rail infrastructure facilities. The Dutch government finances the construction of new infrastructure. User charges: YES 2000?	Punctuality, Safety levels, Expected financial results,	Puts much more in multiyear-contracts with the railway undertakings.	Access rights linked to the public services contracts, 10 years.	One or more regional railway undertaking other than the incumbent operator. Experience with tendering for PSO contracts. Budget responsibility at regional level.
Norway	Infrastructure: Public Operations: Public						
Portugal	Infrastructure: Public Operations: Public		The REFER is responsible for the infrastructure area.	Number of passengers, Safety levels Expected financial results. Statistical procedures in Portugal less reliable than in some other EU countries.	Puts PSO in the law or secondary legislation and compensation is established year by year. No public service contract does exist.	Unlimited exclusive access rights	One or more regional railway undertaking other than the incumbent operator. Experience with tendering for PSO contracts.

Table A7: Part one (continued)

	Ownership type	Financial rehabilitation	Infrastructure	Performance criteria (for public authorities)	PSO in the national legal context	Access rights	Regional rail transport
Spain	Infrastructure: Public Operations: Public		The infrastructure is owned and managed by the associated operating companies.	Safety levels, Financial results.	Puts much more in multiyear-contracts with the railway undertakings.		One or more regional contract with regional public authorities carried out by the incumbent operator. Budget responsibility at regional level.
Sweden	Infrastructure: Public Operations: Public		Managed by a public agency (BV).	Performance of BV is monitored by the government.			
Switzerland	Infrastructure: Public + private Operations: Public + private	in 1998?	Infrastructure is owned by the railway companies. User charges:				
United Kingdom	Infrastructure: Private Operations: Private		Infrastructure providers in the UK come under pressure from different sources: Railtrack must balance calls from the ORR and and the TOCs for new investment, with its financial obligations to shareholders.	PassKm, Punctuality, Expected financial results, Safety levels	Puts much more in multiyear-contracts with the railway undertakings.	Access rights linked to the public services contracts, 7-15 years.	One or more regional railway undertaking other than the incumbent operator. Experience with tendering for PSO contracts. Budget responsibility at regional level.

Table A7: Part two

	Entry Regulation	Price Regulation	Qualitative Regulation	Quality indicators	Quantity of services	Integration between services	Financial compensation and Incentives	Remedies for passengers
Austria	**NO**	**NO** Regular fares can freely be set by OBB. The contract foresees (minimum) substantial general tariff reductions and reductions for special groups and also for weekly/monthly tickets.	**YES** How? Four year contract regulation between OBB and the Federal state, but the specific level of quality is defined year by year.	All forms of payment: credit cards, cash, checks). Customer satisfaction survey, Travel information (Regional transport), Quality of the Rolling stock, to passengers: seat quality, seat availability, cleanliness, and to environment: diesel, electrical. Accessibility of disabled people .	Minimum amount of TrainKm	Ticket integration with services run by local authorities.	Extra payment for more train-Km.	Up to decentralized authorities.
Belgium	**YES** How? Rolling stock and train personnel must be approved by the NMBS.	**YES** How? Basic prices are regulated by contract between the government and NMBS. NMBS has the right to propose other commercial formulas (concerning elderly or disabled people).	**YES** How? Responsibility of the Minister of Transport to provide railway passenger services according to contracts between the state and the railway operators. Safety checks, working and operating safeguards and guidelines.	Information to passengers: timetables, price, general conditions of services, schedules services (delays...). Maintenance of stations: accessibility (especially for disabled people), ticket offices, platforms. Parkings: increase of the number of places. Comfort of travel: seats, air conditioning, facilities for disabled people... Punctuality. Customer service. Accessibility of disabled people. Complementarity with the other means of public transport. Complementarity train/bike. Customer compensation system.	Minimum level of services expressed in the form of trains-Km and in the form of trains per day.	Duty of coordination of timetables between SNCB and other public transport operators (TEC, DE LIJN, STIB). Ticket integration (mainly in Brussels).	Management contract: minimum level of service and associated financial compensation from the government for offering these services and investment in infrastructure and rolling stock. State contributions linked to the increase in the number of passengers. Possibility of increasing tariffs in case of good punctuality performances	

Table A7: Part two (continued)

	Entry Regulation	Price Regulation	Qualitative Regulation	Quality indicators	Quantity of services	Integration between services	Financial compensation and Incentives	Remedies for passengers
Denmark	**YES**	**YES** How? Decided by DSB (fares cannot be increased more than the annual inflation). DSB must give discounts to elderly people, childs…	**YES** How? Safety checks, working and operating safeguards and guidelines. The performances of the Finnish Rail Administration are monitored.	Access to information about timetables and ticketing.	Minimum amount of train stops per station.	Arrangements to facilitate the use of bus and train in combination.	Penalties to be paid to the Ministry of Transport in case of delayed or cancelled services.	Up to decentralized authorities.
Finland	**YES**	**NO** VR sets all fares (including those for special groups).	**YES** How? The types of trains are determined in the contract. The Ministry has approved the quality of the service and has found it to be on an appropriate level. VR has on its own initiative developed the rolling stock and taken opinions and needs of different user groups into account.	Punctuality. Customer satisfaction: survey	Changes must be approved by the Ministry (VR). Agreements on rolling stock used/ purchased.	Bus services are integrated with rail services. Integrated ticketing by agreement.	For YTV the contract includes a quality sanction/bonus system.	Railway undertaking legally liable for delays.

Table A7: Part two (continued)

	Entry Regulation	Price Regulation	Qualitative Regulation	Quality indicators	Quantity of services	Integration between services	Financial compensation and Incentives	Remedies for passengers
France	**YES** How? Entry regulation is mainly to be differentiated between infrastructure (entry barriers) and operations (entry barriers maintained). Access rights to railway undertakings engaged in the international combined transport or to international groupings when one of the constituting company is established in France for the international transport services.	**YES** How? Fixed by SNCF with the agreement of the State. Social tariffs for special groups of passengers. Regions may establish special tariffs compatible with the general tariffs of SNCF.	**YES** How? At the national level, SNCF must guarantee the best conditions of safety, accessibility, comfort and punctuality. At the regional level, SNCF and the Region shall define the level of quality (continuity of service, information to passengers, complaints handling) to be reached and they will formalize it in the agreement.	Safety, Comfort, Punctuality information, Accessibility.	A sufficient number of seats must be guaranteed.	SNCF shall facilitate connections with other means of transport. The Regions shall guarantee integrated tariff schemes.	Based on punctuality and the quality of service.	Railway undertaking itself pays compensation to passengers unless force majeure.
Germany		**NO** The regular fare levels are not regulated. Fare reduction for long distance and regional transport. Free transport to severely handicapped persons.	**NO** No regulation at the national level.		Obligation to transport and obligation to operate services.			Up to decentralized authorities.

Table A7: Part two (continued)

	Entry Regulation	Price Regulation	Qualitative Regulation	Quality indicators	Quantity of services	Integration between services	Financial compensation and Incentives	Remedies for passengers
Greece								
Italy		**YES/NO?** How? Ticket facilities are granted to young people, the elderly and disabled persons and to figures belonging to special administrations.					Ticket prices are linked to the achievement of quality parameters.	Up to decentralized authorities.
Ireland								
Luxembourg		**YES** How? Defined by the Ministry of Transports. CFL is forced to adopt these fares but it can propose an adaptation to take into account increased costs.	**YES** How? The Ministry of Transports defines the quality of service. CFL is bound to run the services with a particular regard to its quality.	Punctuality, Access to stations, Comfort of the journey.	Frequency of trains and stations to be served	CFL shall take into account the integration between railway transport and other forms of public transport.		
The Netherlands	**YES** How? Access to the passenger market is conditional upon the subsidy requirement and the desirability of the service.	**YES** How? General tariff obligation on NS not to increase tariffs for selected fares (single, return, monthly tickets) more than 2% above the consumer price index per year.	**YES** How? In the performance contract, quality of the rail passengers' services is contracted. Safety checks, working and operating safeguards and guidelines.	Punctuality, Accessibility for people with reduced mobility, Financial compensation for individual passengers with delays of more than 30 minutes.	Minimum frequency, first and last train.	Integration of timetables, ticket integration and information integration	Financial penalty if punctuality performance is below a certain level. Financial bonus if the number of passengers is above a certain level. Compensation to passengers.	Railway undertaking obliged in PSO contract to pay compensation to passengers.

Table A7: Part two (continued)

	Entry Regulation	Price Regulation	Qualitative Regulation	Quality indicators	Quantity of services	Integration between services	Financial compensation and Incentives	Remedies for passengers
Norway	**YES** How?	**NO**	**YES** How?					
Portugal		**NO** Fares may be set freely for journeys longer than 50 km. Reductions for elderly people, children, disabled persons and pensioners.	**YES** How? Agreement between SNRIPD, REFER and CP to promote accessibility on the railways. Safety checks, working and operating safeguards and guidelines.	Environmental characteristics of rolling stock, Complaints from passengers, Punctuality, Accessibility.	Minimum number of trains. Agreements on rolling stock used / purchased.	Agreements between operators for ticket integration and integration between different modes of transport.	Based on punctuality and regularity of services (Fertagus).	Up to decentralized authorities.
Spain		**YES** How? Prices are regulated for social routes and deregulated for commercial services.	**YES** How? RENFE public service obligations are mainly defined in a pluriannual contract between RENFE and the state. The quality indicators are defined in the contract.	Punctuality, Quality of tracks, Number of accidents.	Trains per Km and passengers per Km are defined in the contract (RENFE).	Functional integration between RENFE and FEVE in the Metro/Train of the Asturias.		Up to decentralized authorities.
Sweden	**YES** How? The market is opened to foreign competition on long distance transports.	**NO**	**YES** How? Safety checks, working and operating safeguards and guidelines.					

Table A7: Part two (continued)

	Entry Regulation	Price Regulation	Qualitative Regulation	Quality indicators	Quantity of services	Integration between services	Financial compensation and Incentives	Remedies for passengers
Switzerland	**YES** How? Third party access will be granted after a approval procedure at the Federal Office of transports.	**YES** How? Prices for passengers transport are monitored.						
United Kingdom	**YES** How? The ORR regulates Railtrack according to RPI-x formula, whilst, OPRAF monitors and enforces the franchise contracts signed by the 25 TOCs.	**YES** How? If necessary, the franchising director must include in the relevant franchise agreement a provision for securing that the prices or fares are reasonable. Mandatory Card Schemes (Young Persons Railcard, Senior Railcard, Disabled Persons Railcard) and Voluntary Card Schemes (Family and HM Forces Railcards and Network Card).	**YES** How? The franchise agreement requires a franchisee to establish its own passenger's charter which must include target for some quality indicators. Safety checks, working and operating safeguards and guidelines.	Punctuality, Reliability, Queuing times at ticket offices, Information, Compensation arrangements, Accessibility for disabled people.	Minimum passenger rail services, frequency of trains, stations to be served, maximum journey times, first and last trains, weekend services, through services, seats availability, peak train capacity.	In Northern Ireland, NITHC is introducing an integrated ticketing/fares regime to facilitate intermodal travel (buses and trains).	Compensation to passengers.	Railway undertaking obliged in PSO contract to pay compensation to passengers.

REFERENCES

Affuso, L., and D. Newbery (2000): "Investment, Reprocurement and Franchise Contract Length in the British Railway Industry," Discussion Paper 2619, CEPR, London.

Aghion, P., and P. Howitt (1998): *Endogenous Growth Theory.* Cambridge, MA: MIT Press.

Amici della terra, Ferrovie dello Stato (2002): *The Environmental and Social Costs of Mobility in Italy*. Fourth Report.

Baily, M., and H. Gersbach (1995): "Efficiency in Manufacturing and the Need for Global Competition," *Brookings Papers on Economic Activity* : *Microeconomics*, p 307-347.

Bain, J. (1959): *Industrial Organization*. New York : John Wiley.

Blundell, R., R. Griffith, and J. Van Reenen (1995): "Dynamic Count Data Models of Technological Innovation," *Economic Journal*, 105:429, p 333-344.

Bolton, P., J.F. Brodley, and M.H. Riordan (1999): "Predation Pricing, Strategic Theory and Legal Policy," Discussion Paper 9982, Tilburg Center for Economic Research.

Braeutigam, R. (1979): "Optimal Pricing with Intermodal Competition," *American Economic Review*, 69:1, p 38-49.

Braeutigam, R. (1999): "Learning about Transportation Cost" in *Essays in Transportation Economics and Policy: A Handbook in Honor of John R. Meyer*. José A. Gómez-Ibáñez, William B. Tye, (eds), Clifford Winston, (ed). The Brookings Institution, Chapter 3, p 57-98.

Bresnahan, T., and P. Reiss (1991): "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99:5, p 977-1009.

Bureau of Transport and Regional Economics. Australia (2002). Data Base.

Campos, J., and P. Cantos (1999): "Rail Transport Regulation," Working Paper, World Bank Series.

Cantos, P. (2001): "Vertical Relationships between Railways Operation and the Infrastructure for the European Case," *Transport Journal.*

Carlin, W., S. Fries, M. Schaffer, and P. Seabright (2001): "Competition and Enterprise Performance in Transition Economies: Evidence from a Cross-country Survey," Working Paper 376, William Davidson Institute.

Carlin, W., J. Haskel, and P. Seabright (2001): "Understanding 'the Essential Fact about Capitalism': Markets, Competition and Creative Destruction," *National Institute Economic Review*, 175, p 67-84.

Department of Transport of U.K. (2001). Transport Statistics Edition.

ECMT (1998): "Efficient Transport for Europe, Policies for Internalisation of External Costs". Paris: OECD.

European Commission (2001): *Directive 2001/14/CE of the European Parliament and of the Council of 26 february 2001 on the Allocation of Railway Infrastructure Capacity and the levying of Charges for the Use of Railway Infrastructure and Safety Allocation* (OJL075, 15/03/01).

European Commission. DG Energy and Transport Statistics (2002).

Friedriech, R., and P. Biekel (eds) (2002): *Environmental External Costs of Transport*. Berlin: Springer Verlag.

Greene, D., D. Jones, and M. Delucchi (eds) (1997): *The Full Costs and Benefits of Transportation*. Berlin: Springer Verlag.

Hayashi, Y., K. Button, and P. Nijkamp (eds) (1999): *The Environment and Transport*. Cheltenham, UK; Northampton, Mass.: Edward Elgar.

Infras and Iww. (2000): "External Costs of Transport: Accident, Environmental and Congestion Costs in Western Europe".

Ivaldi, M., and G. McCullough (2001): "Density and Integration Effects on Class I Freight Railroads," *Journal of Regulatory Economics*, 19:2, p 161-182.

Ivaldi, M., and G. McCullough (2002): "Subadditivity Tests for Network Separation using a Generalized McFadden Cost Function," IDEI Working Paper.

Jones-Lee, M.W. (1989): *The Value of Safety and Physical Risk*. Oxford: Basil Blackwell.

Jones-Lee, M.W., and G. Loomes (1994): "Towards a Willingness-to-Pay Based Value of Underground Safety," *Journal of Transport Economics and Policy*, 28, p 83-98.

Kolm, S. (1968): "Prix publics optimaux," *Cahiers du séminaire d'économétrie*, n°9, Paris.

Kwoka, J.E. (1996): *Power Structure: Ownership, Integration, and Competition in the U.S. Electricity Industry*. Boston; Dordecht and London : Kluwer Academic.

Levy-Lambert, H. (1968): "Tarification des services à qualité variable-Application aux péages de circulation," *Econometrica*, 36:3-4, p 564-574.

Litman, T. (1999): *Transportation Cost Analysis : Techniques, Estimates and Implications*, online edition, Victoria Transport Policy Institute (Source: http://www.vtpi.org/tca).

Litman, T. (2002): "Evaluating Transportation Land Use Impacts," Victoria Transport Policy Institute.

Lundberg, A. (1996): "Restructuring of Swedish State Railways," *Japan Railway and Transport Review*, n°8, September 1996, p 22-26.

Mauch, H., and W. Rottengather (1996): "External Cost of Land Transport," Report to UIC.

McFadden, D. (1977): "The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban transportation" in *Emerging Transportation Planning Methods*, U.S. Department of Transportation DOT-RSPA-DPB-50-78-2.

Megginson, W., and J. Netter (2001): "From State to Market : A Survey of Empirical Studies on Privatization," *Journal of Economic Literature*, 39:2, p 321-389.

Millward, R. (2002): “European Governments and Infrastructure at the Start of the 20th Century : Economics. Ideology, Defence,” University of Manchester, mimeo.

Mizutani, F., and K. Nakamura (2001): “The Japanese Experience with Railway Restructuring,” Kobe University, mimeo.

Mizutani, F., and K. Shoji (2001): “Operation-Infrastructure Separation in the Japanese Rail Industry: The case of Kobe Kosoku Testudo,” Kobe University, mimeo.

Morrison, S.A., and C. Winston (1986): *The Economics of Airline Deregulation*, Washington, D.C.: The Brookings Institution.

Nash, C., and J.P. Toner (1998): Background Note for OECD Round Table.

Nellthorp, J., P. Mackie, and A. Bristow (1998): “Measurement and Valuation of the Impacts of Transport Initiatives. Deliverable D9”, (Restricted), EUNET Project - Socio Economic and Spatial Impacts of Transport (Contract: ST-96-SC.037), Institute for Transport Studies, University of Leeds.

NERA (1998): An Examination on Rail Infrastructure Charges. *Report for European Commision DG-VII.*

NERA (2001): *Global Antitrust Weekly*, March 2001, from n°121 to n°125.

Neven, D.J., and L.H. Röller (1996): “Rent Sharing in the European Airline Industry,” *European Economic Review*, 40:3-5, p 933-940.

Ng, C.K., and P. Seabright (2001): “Competition, Privatization and Productive Efficiency: Evidence from the Airline Industry,” *Economic Journal*, 111:473, p 591-619.

Nickell, S.J. (1996): “Competition and Corporate Performance,” *Journal of Political Economy*, 104, p 724-746.

Nicoletti, G., and S. Scarpetta (2002): “Regulation, Productivity and Growth: OECD Evidence,” Economic Policy, n° 36, forthcoming, available from http:www.economic-policy.org.

OECD (1988): *Transport and the Environment*. Paris: OECD.

OECD (1998): “Railways: Structure, Regulation and Competition Policy. DAFFE/CLP(98)1, Les tables rondes sur la politique de la concurrence, n°15, ” OECD.

Oum, T.H., W.G. Waters II, and J.S. Yong (1990): “A Survey of Recent Estimates of Price Elasticities of Demand for Transport”. Working Paper WPS 359, World Bank.

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, p 9-42.

Pearce, D. (1993): *Economic Values and the Natural World*. Cambridge, MA: MIT Press.

Peltzman, S., and C. Winston (2000): “Deregulation of Network Industries: What's next?” Washington D.C.: AEI-Brookings Joint Center for Regulatory Studies.

Preston, J. (2001). “Regulation Policy in Land Passenger Transportation in Europe,” presented to the 7th International Conference on Competition and Ownership in Land Passenger Transport, Molde, Norway, 25-28 June 2001.

Productivity Commission Report (1999): "Progress in Rail Reform", Australia.

Quinet, E. (1994): "The Social Costs of Transport: Evaluation and Links with Internalisation Policies" in *Internalizing the social costs of transport*, Paris: EMCT, OECD.

Rey, P., P. Seabright, and J. Tirole (2001): "The Activities of a Monopoly Firm in Adjacent Competitive Markets: Economic Consequences and Implications for Competition Policy," IDEI Working Paper.

Scherer, F.M (1980): *Industrial market structure and economic performance*. Boston: Houghton Mifflin.

Shires, J., and J. Preston (1999a): "Getting Back-on Track or going off the Rails? An Assessment of Ownership and Organizational Reform of Railways in Europe." To be presented to the Sixth International Conference on Competition and Ownership in Land Passenger Transport, Cape Town, South Africa, 20-23 September 1999.

Shires, J., J. Preston, C. Borgnolo, and M. Ponti (1999b): "SORT IT Deliverable VI. Joint Report on the Rail Case Study." European Commission, Brussels.

SIKA (1999): *Översyn av samhällsekonomiska kalkylprinciper och kalkylvärden på transportområdet. ASEK. Redovisning av regeringsuppdrag, juni 1999*. Stockholm: Statens Institut för kommunikationsAnalys, SIKA Rapport 1999:6.

Small, K.A. (1992): *Urban transportation economics*. Chur (Switzerland): Harwood Academic Publishers.

SJ –Swedish State Railways, Annual Report 2001, Sweden.

Tervonen, J. (1999): *Accident costing using transfer values*, VTT Technical Research Center of Finland, Espoo Finland.

Thompson, L.S. (1997), "World Bank Support for Developing Railways of the World," *Japan Railway & Transport Review*.

Union Internationale des Chemin de Fer (2002). "Quarterly Statistics".

US. Department of Transport. Bureau of Transport Statistics (2002).

Viscusi, W. Kop (1993): "The Value of Risks to Life and Health", *Journal of Economic Literature*, 31, p 1912-1946.

Williamson, O.E. (1985): *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, New York: The Free Press.

Willig, R.D. (1987): "Corporate Governance and Market Structure" in A. Razin and E. Sadka (eds), *Economic Policy in Theory and Practice*, London: Macmillan, p 481-494.

WEBSITES :

http://www.jica.go.jp/english/publication/network/net_vol14/03main03.html
http://www.sncf.com/co/rapport99/PAGE/ang/strat.html
http://www.cenorm.be/sectors/transport/intertrans/resources/revprog.pdf
http://www.sice.oas.org/FTAA/cartage/sumcon/workse7.asp
http://people.hofstra.edu/geotrans/eng/ch3en/conc3en/ch3c5en.html
http://www.princeton.com/intermodal/
http://www.business.com/search/rslt_default.asp?query=intermodal+passengers&bdcf=F679364A-DD99-11D3-82D2-00C04F01021C&bdcs=EF80DFB7-E2B0-4754-8FC0-BCCD9F7C0297&search=search&vt=all
http://www.cfr.ro/JF/engleza/nr5/railways_21.htm
http://www.cordis.lu/transport/src/taskforce/src/act6a.htm
http://europa.eu.int/scadplus/leg/en/lvb/l24094.htm

http://www.dotars.gov.au/btre/index.htm
http://www.uic.asso.fr/d_stats/stats_en.html
http://europa.eu.int/pol/trans/index_en.htm
http://www.transtats.bts.gov/
http://www.publicpurpose.com/tfb-ix.htm