RESTRUCTURING RAILWAYS: IN THEORY AND PRACTICE

MARC IVALDI*

December 2005

Introduction

After years of permanent decline, the European railways industry is experiencing changes. A clear sign of these changes is found in the coexistence of the old and the new at the present time, which yields a highly contrasted picture. Here the passenger is wondering if the nightmarish train in which he/she finally finds a relatively clean seat belongs to the same world of this millennium; there he/she is cocooned while moved at 300 kph (and soon even more) on a high speed train between Paris and Brussels. Whether these changes are significant, and whether the policies underlying them are relevant for this industry to take on a new lease of life, to be part of a sustainable growth in Europe and, in some ways, to recover the glory attached to its role in the first half of the 20th century, motivates a lively debate in Europe and elsewhere.

This article is aimed at providing some light on some of the main economic issues that are on the agenda of this evolving industry.[1]

The last three decades have seen a dramatic decline of EU railways while the transport industry in Europe as a whole has grown steadily at a 3% average rate. Rail freight transport has lost two third of its market share to the benefit of road and maritime freight transport. Over the period 1970-2000, its market share has decreased to 8% from 21%. Although of a smaller size, the loss is also significant for rail passenger transport whose market share has vanished to 6% from 10%.[2]

The European case contrasts with the state of the US railways industry. The restructuring of US freight markets undertaken in the early eighties has allowed rail industry to dramatically improve its efficiency—reducing both its workforce and its track-mileage by 50 percent. In the face of competition between truck and rail and between integrated firms, the rail freight industry has maintained its level of activity by focusing especially on bulk and intermodal shipments.[3] The rail share of the freight market on a tonnage basis was 11.8 percent (compared to 54.5 percent for truck) in 1993 and 12.2 percent (compared to 58.2 percent for truck) in 2002. It is also important to note that during the period 1980-2005, the US

* Marc Ivaldi is Professor of Economics at University of Toulouse and EHESS, Researcher at Institut D'Economie Industrielle (IDEI) and Industrial Organization Programme Co-Director at CEPR.

[1] This article draws heavily from a report on the economics of railroads made by a team of IDEI researchers led by Paul Seabright. See http://idei.fr/doc/wp/2003/rapport_db_1.pdf.

[2] These figures are quoted by Di Pietrantonio and Pelkmans (2004) and also by Nash, Matthews and Shires (2004).

[3] Note that in the U.S., railroads own their tracks, they compete for shipments and they rent to competitors the access to their tracks.

rail industry experienced a significant merger wave and achieved a high level of concentration. Despite this dramatic industry consolidation, the consumer surplus in U.S. rail freight markets has increased by about 30 per cent between 1986 and 2001, suggesting that to date the trade-off between merger-specific efficiency gains and merger-related increased market power has favoured rail customers. (See Ivaldi and McCullough, 2005.)

In worldwide perspective however, the railway industry faces strong challenges from other modes and complex problems. All attempts to restructure this industry focus on two main questions: What is the best structure for this industry? Is competition feasible and what are the conditions for efficient entry? We provide some replies to these questions.

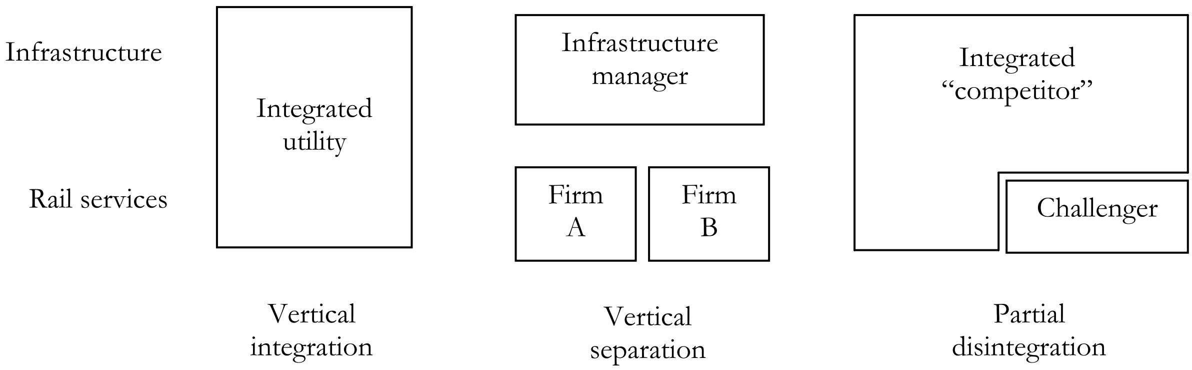
Vertical disintegration

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.[4] 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 favour 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 1).

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 trickier, 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 and the public authority faces a complex problem of combined sectoral and competition regulation. (See Rey *et al.*, 2001)

[4] 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.

Figure 1: Alternative vertical organizations



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 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 complementarities in freight transport between infrastructure and operations for USA railroads using a translog specification. According to the estimates shown in Table 1 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 1: 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[5] 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 2 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.[6] 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.

Table 2: Cross-Elasticities of Marginal Operating Costs with respect to Infrastructure

		CROSS ELASTICITIES	
Firm		Passenger Marginal Cost with respect to infrastructure	Freight Marginal Cost with 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).

[5] The monetary value of all infrastructure facilities (track, buildings, stations, etc) is employed as a variable for measuring "Infrastructure Output".
[6] As reported in Mizutani and Nakamura (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 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 show that vertical disintegration of infrastructure from operations could represent a significant cost penalty so that it is wise to account for them if one decides to disintegrate the system, at least to compensate for transaction costs. However, given that such separation has occurred very rarely to date or more precisely is too recent, the value of such studies in predicting the future consequences of such separations is still limited.

Such studies cannot realistically shed light on one important issue which 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.

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 again later.

Secondly, both entrants and incumbents 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.[7]

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.

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:

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

[7] On these points see Antes *et al.* (2004) and Ivaldi and Vibes (2005).

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.[8] 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.

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.

Then what is the likely outcome of on-track competition? 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.

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 $u_e - p_e \geq \overline{u}$). 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 $c_i \leq c_e$.For frequent travelers, a

8 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.

tariff such that $p_i \leq u_i - u_e + c_e$ prevents any entry. Even if $c_i > c_e$, the switching cost $u_i - u_e$, which is positive, 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).

The nature of competition is also determined by the nature of the costs faced by the entrant and the incumbent. This affects 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,[9] plus an additional element, i.e., the opportunity cost, which is any additional net cost incurred on those connecting routes to which some of the passengers may subsequently transfer.[10] When transferring passengers in fact yield a profit on the connecting routes, this opportunity cost is negative, and an incumbent's true cost lies below a conventionally-measured accounting measure of its costs of providing the service. Three consequences follow from this:

First, in networks where connecting traffic is a comparatively large fraction of overall traffic (like Germany but unlike France, say), there are fewer cherries for entrants to pick, i.e., there are relatively few connections with a high point-to-point demand.

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 is usually lower than conventional cost comparisons would lead us to predict.

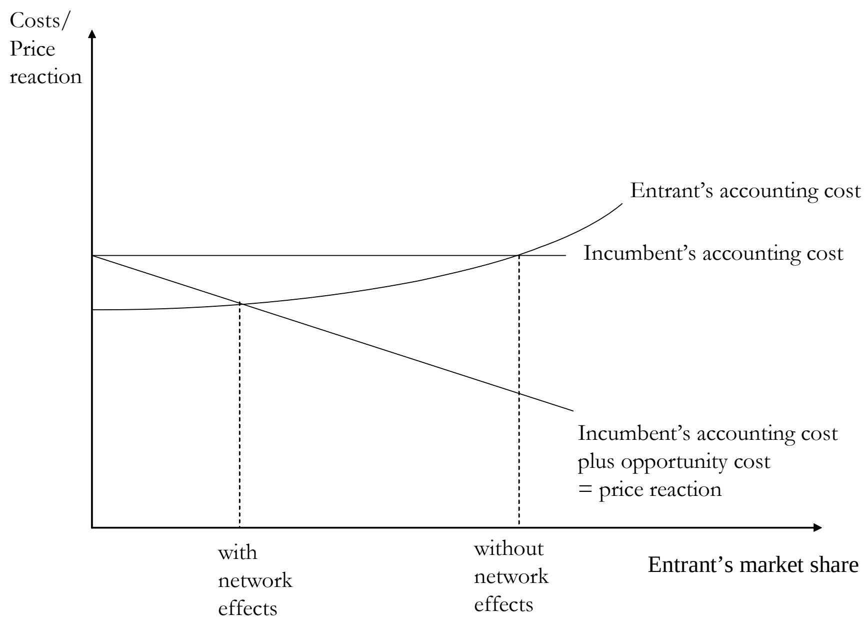
Thirdly, it is likely that the Mohring effect (see Small 1992) means that the opportunity cost element becomes more important as the entrant's market share increases, since the reduction in the value of frequency of service to passengers becomes progressively more important as the frequency itself

[9] 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.

[10] 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.

declines.[11] 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 that not only will the incumbent's true cost lie below its accounting cost, but it will be significantly more steeply sloped. (See Figure 2.) Therefore even an entrant with a significant initial cost advantage finds 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. 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 2: Marginal passenger costs at different levels of entrant market share



[11] The Mohring effect is an external effect stemming from the increase of traffic in public transport. This effect lead to a positive externality: When traffic increases, the operator is led to increase the frequency of the services, thus by reducing the waiting time it reduces the total travel time of the users.

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 leads to large market shares for the entrant. This is not, however, to say that we can predict exactly how large such market shares is, since circumstances varies significantly from route to route. Secondly, the share of the incumbent is 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 goes beyond the scope of this article.

Concluding remark

Here many questions have not been discussed like competition for the track, the role of regulation, the choice of investments, the structure of the industry producing equipments for railways among many others. The economics of railways is a promising domain for research in industrial organization.

References

Antes, J., G. Friebel, M. Niffka and D. Rompf (2004), "Entry of Low-Cost Airlines in Germany. Some Lessons for the Economics of Railroads and Inter-Modal Competition", mimeo, Deutsche Bahn and University of Toulouse.

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

Di Pietrantonio, L. and J. Pelkmans (2004), "The Economics of EU Railway Reform", *Bruges European Economic Policy Briefings*, no. 8.

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.

Ivaldi, M., and G. McCullough (2005), "Welfare Tradeoffs in U.S. Rail Mergers," IDEI Working Paper.

Ivaldi, M., and C. Vibes (2005), "Intermodal and Intramodal Competition in the Passenger Transport Market,", IDEI Working Paper.

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.

Nash, C., B. Matthews and J. Shires (2004), "The EU Transport Policy White Paper: An Assessment of Progress", Institute for Transport Studies, University of Leeds.

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.

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

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

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