7TH WCTR
SYDNEY AUSTRALIA
1 9 9 5

TOPIC 13
PUBLIC SECTOR
PERFORMANCE

SWEDISH RAIL POLICY: A CRITICAL REVIEW

NILS BRUZELIUS
University of Växjö
Dept. of Transport Economics and Logistics
Växjö, SWEDEN

ARNE JENSEN
University of Gothenburg
Dept. of Business Administration
Göteborg, SWEDEN

LARS SJÖSTEDT
Chalmers University of Technology
Dept. of Transportation and Logistics
Göteborg, SWEDEN

Abstract

An overview of the reform process is given. A microanalytic study evaluates competitive pressure and potential for increased efficiency of the present system. For some markets economies of scale with and without internal competition are estimated using real cost data. Existence of market failures is discussed and perception of railways as natural monopolies is challenged.

INTRODUCTION

Background

In terms of the Swedish railway model, the State assumes direct responsibility for constructing and maintaining the infrastructure which at a charge is put at the disposal of the railway operators. This split of responsibility for the railway system has, with the creation of the Swedish National Rail Administration (BV) 1 January 1989, in Sweden gone further than anywhere else in the world. Provided that this split of responsibility is appropriate both in a short and long-term perspective, the logical next step would be to let more than one operator use the railway infrastructure. Since competition is considered highly beneficial to efficiency, governments normally make sure, that commercial activities are structured in such a way that consumers may choose between several suppliers of similar products. According to the transport policy decision of 1988, the Swedish State Railways (SJ) has the right to operate freight services on the entire Swedish railway network and passenger services on the main lines. The concessions for passenger service on the county lines have been transferred to the County Transport Authorities. It is now being proposed to allow other operators to supply the same kind of services for which SJ today has the exclusive right—ie to allow intra-sector competition, here called *internal competition*, on the Swedish railway network. In principle two forms of internal competition on a given line, for a given service are conceivable: *competition for the market* and *competition in the market*. In competition for the market, an operator receives after a call for tenders, the sole right for the service within a certain period whereas, what is here called competition in the market, is based on two or more operators being allowed to operate in parallel on the line for the demand to which their services is directed.

Objectives of paper

The issue inspiring this research work was: "Will competition between several enterprises on the same track lead to more effective railway service for the country as a whole?" Since a negative answer will lead to questioning the Swedish model as a whole it is, however, logical to widen the question as follows: "Will the Swedish railway system in its entirety be more effective if the State is given responsibility for the track, while one or more enterprises handles the operations, as compared to giving full responsibility for both track and operations to a single enterprise?"

The Swedish reform process

SJ was founded on 1 December 1856. Today, the whole network is nationalised. BV manages and maintains a network of main lines (6,239 km), the Ore Line (744 km) and county lines (3,325 km). SJ was from the beginning a *public enterprise*. This form of organisation harmonises with the Swedish public administration tradition, according to which authorities and agencies are independent inasmuch as the Government only prepares the rules, instructions, and budget, which are then submitted to the Parliament for approval. Before the split of 'old' SJ into 'new' SJ and BV the only but important exception to the public enterprise principle was the National Road Administration. The main reasons for not attempting to run the National Road Administration as a public enterprise were the difficulty to put charges on its services, and the need for local political influence on physical design and use. Instead of defining a State road capital on which the National Road Administration would have to pay interest, the principle has been to write off all the National Road Administration's investments immediately. In return, all those who use the road network had to pay vehicle and fuel taxes directly to the public treasury.

A big step towards deregulation was taken in 1963 when SJ first was allowed to conclude non-public freight agreements with customers and simultaneously underwent a financial reconstruction. The main principle was that each transport mode should cover its own costs. Rail lines carrying little traffic that not yet had been closed down were transferred to the 'Weak Traffic

Network' entitled to financial compensation. The cost responsibility for this network was taken over by State, which paid for a negotiated supply of transport services through what came to be known as the 'collective ticket'. The 1963 transport policy decision led to an intense debate in the 1970s. A number of economists asserted emphatically that SJ's prices must rest on socio-economic principles and obey the marginal cost principle. In contrast SJ's management at the time firmly maintained that a general acceptance of the marginal cost principle would prevent necessary railway closures and, in practice, entail that it would not be possible to cover the costs of investments necessary to renew railway infrastructure and rolling stock. SJ lost the battle. In terms of the 1979 transport policy decision, Parliament confirmed the marginal cost principle. Steps were taken to restructure SJ's finances but the fundamental problem of coping with the rapidly growing lack of funds for maintaining existing infrastructure, and for the increasingly acute needs of renewal was not properly solved. As a consequence of the unsolved structural problems SJ's losses, after a few years, once again rose steeply. It gradually became obvious that railway policy would once again have to be reformed.

The basic idea of the reform is to apply the so called *road transport model* to the railway system. By separating infrastructure installations from SJ and transferring them to an agency emulating the National Road Administration, SJ would be relieved of its century old obligations of covering the full costs of all railway infrastructure including depreciation and nominal rate of return on the state capital accumulated in this infrastructure. Several positive effects were expected as a result:

- The conflict of objectives between production of infrastructure and transport services would be solved by letting socio-economic principles guide the planning and pricing of infrastructure services, while maintaining public control and complete transparency. Simultaneously SJ is given freedom to apply a market approach based on normal business practises, and no longer required to make publicly available other economic and statistical data than private enterprises.
- By relieving SJ from most of the infrastructure costs long-term conditions would be created that would allow SJ to generate a sustained profit. The political justification of this move was that the low marginal costs, and the high degree of safety and environmental friendliness of railway transport motivate low infrastructure charges.
- The railway's image would be improved by making means which had earlier been viewed as subsidies look like future-oriented investments in and maintenance of infrastructure.
- Decision processes of investments in infrastructure would be facilitated, through the use of social cost-benefit analysis.
- By cutting right through SJ's organisation favourable starting-points for a renewal process would be created.

The border line established between BV and 'new' SJ largely follows the common notion of what belongs and does not belong to infrastructure. Major marshalling yards were allocated to BV whereas minor ones stayed with SJ. The most complex question was that of traffic control. In SJ's earlier internal organisation model traffic control was consistently classified as part of the infrastructure. Accordingly, in the 1987/88 Government Bill it was proposed that responsibility for traffic control should be given to BV. Realising the importance of maintaining full control of its operations SJ fought for keeping its right to exercise the actual operation of the traffic control system.

It was finally decided to transfer only the equipment used for traffic control to BV, whereas both cost and operating responsibility for traffic control was left with SJ. A financing agreement was concluded. between BV and SJ, in terms of which both are responsible for planning and implementation whereas SJ is responsible for transport production. The main motive appears to have been that immediate traffic control should be placed within or as close to transport production as possible.

As a consequence of the split, BV and SJ have signed a general agreement regulating BV's track access for maintenance purposes. This agreement stipulates that individual agreements shall be negotiated and signed annually at the regional level for each railway line. These are supplemented with short-time agreements covering the next few weeks which are updated every two weeks. An important part of the annual agreement is the reliability of each line. The goals for this have been set through mutual understanding. SJ undertakes to limit constraints as to access time to different kinds of track to a maximum number of minutes a week whereas, on the other hand, BV guarantees that track malfunctions will not exceed a given number over a period of one year. BV also guarantees that reporting for duty to carry out instant repairs will take place within a given time after a failure being communicated.

The road transport model implies that BV operates in terms of the same principles as the National Road Administration. Parliament then appropriates funds for both investments and maintenance. The procedure implies that infrastructure investments are written off immediately. In order to place SJ on an equal footing with owners of vehicles using roads, track charges are imposed. These are meant to correspond to road taxes for vehicles and fuel. For each railway vehicle registered for transport use a track charge is imposed according to the formula: $Y_i = a_i + b_i \cdot X_i$, where Y_i is the total annual charge, X_i the vehicle use expressed in gross tonne-kilometres per annum and a_i and b_i the coefficients for the fixed and variable charges respectively regarding the type of vehicle i. The price coefficient b_i is meant to correspond to the short-term marginal socio-economic costs which the vehicle causes on different types of tracks. The fixed charge a_i is meant to be chosen in order to bring about cost neutrality in relation to road transports. Railway wagons being used in intermodal transport are exempt from the fixed charge. However, the level of the charges are fixed by cabinet as it sees fit. Revenues generated by the track changes are for the treasury and are not ear marked.

Results of the reform

A deeper analysis of the results of the reform has not yet been presented. It is obvious that it has on several levels brought about big changes. BV and SJ have both very quickly presented detailed investment plans and a number of extensive infrastructure projects have been completed or started. A question that has only recently been brought into the general debate is, if it is reasonable to finance such a large part of all investments in the railway sector by public funds. For the fiscal year 1992/93 the Government has given BV permission to use, in total, the amount of SEK 8,170 million. Track charges paid to the State treasury in the calendar year 1992 only amounted to SEK 692 million.

THE UNIQUE PROPERTIES OF RAILWAY OPERATIONS

In the remainder of this paper it will be argued that the Swedish model is flawed. The basic reason is that railway operations cannot be compared with road operations, ultimately on account of the special properties of the railway technique.

A road vehicle has two degrees of freedom which in principle allows it to pass other vehicles anywhere in a road network. When congestion develops each additional car driver certainly marginally prolongs driving-time for all others and must himself expect a considerably prolonged travelling time but there are still no restrictions on the availability of the network. The car driver has freedom to start the journey exactly whenever he or she wants to. Therefore market imperfections are normally insignificant.

For railway vehicles the situation is different. They have only one degree of freedom since they are mechanically guided by the rails. Points where trains can change direction certainly exist but because train paths must be locked and protected well before the train passes only one degree of freedom is available to the driver also at points. A train itinerary can only be negotiated if train paths are successively made available to the train as it proceeds. This requires careful preplanning of each train itinerary and means that each local allocation of track capacity to a specific train has global repercussions. To minimise congestion when allocation track capacity and make best use of the whole network a timetable must be produced that covers all prospective train itineraries in the network during a specific time period. These itineraries are the building bricks of the train plan which is central to each operator and inherently determines use of resources and quality of service. Consequently, a transport enterprise must have decisive influence on its train plan and on the timetable of the whole network. Capacity of a railway network can not in a meaningful way be divided into a number of—in relation to one another—clearly and well defined train itineraries which are marketed separately in an 'infrastructure market'.

Structural quality of the timetable is one important element in the success of railway operation. Skill of traffic controllers and responsiveness to operators' needs is the other. It is understandable that operators fear that placing traffic control under autonomous management will impair their quality of service.

COMPETITION ON THE RAILWAY NETWORK

Competitive pressure and efficiency potentials in the railway's transport market

To analyse the degree of competitive pressure in different traffic segments through *external competition*, the simple causal model shown in Figure 1 was used. The railway's *competitive conditions* are such basic, primarily external, factors that the railway can not easily influence and to which it therefore must adjust. Such conditions may either create advantages or disadvantages for the railway in its competition with other modes of transport. *Competitive strategy* is a 3-dimensional concept which determines what transport products the railway should supply to the market, what customers it should market these products to and what production technology it should choose. Competitive strategy constitutes essential dimensions of railway activity in a medium and long-term perspective. *Competitive pressure* influences the production system partly directly, partly indirectly through competitive strategy. Both these effects of competitive pressure are relevant to an analysis of the impact on SJ's efficiency that would result from increased competition. This analytical model differs from traditional economic theory (cf. Leibenstein, 1966 and 1976, and Shepherd, 1979) by giving competitive strategy an explicit role as a possible conveyor of the effect that competitive pressure exercises on efficiency. This approach is influenced by the business economics theories of market-oriented strategic management and summarises aspects treated by Abell (1980), Aaker (1988) and Porter (1980). The model assumes, in other words, that low efficiency may as well be caused by weak strategy as weak implementation.

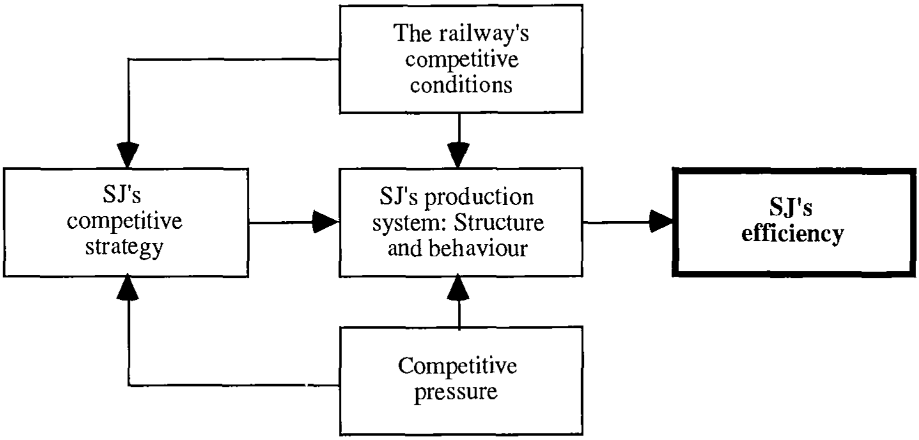

Figure 1 **Analytical model used to find the potential impact on SJ's efficiency from introducing internal competition**

The analysis of competition has primarily identified three segments in which internal competition can be supposed to increase efficiency: domestic combined transport (freight), feeder service in wagon-load transport (freight) and all passenger service bought by society.

Large-scale production advantages and competition in the market

Large-scale production advantages, or shorter *scale advantages*, imply that a given supply can be produced at lower total cost by n producers than by m producers, given that n is less than m. Given that important advantages of large-scale production in railway service would exist for realisable volumes of demand in different market segments in Sweden, two important consequences can be deduced from theory:

- Competition in the market would, other factors being equal, lead to lower socio-economic efficiency due to loss of scale advantages. This is true except for the case where an unused potential for cost rationalisation exists, which may compensate for lost scale advantages but can be realised only by competition in the market.
- It is not likely that establishers will enter the market on strictly commercial, market-economical conditions, if SJ's monopoly is abolished. A condition for entering is that the establisher considers it possible to get down relatively quickly to a unit cost level which does not imply any disadvantage compared to that of SJ. So, the establisher must comparatively swiftly be able to neutralise SJ's volume advantages in the service exposed to competition by gaining market shares, at the same time as SJ's advantages of large-scale production in other services must not be too extensive. This can only be achieved by an establisher, who is both financially perseverant and in possession of a new cost-effective technology which SJ can not immediately copy.

During the last two decades several econometric analyses of railway cost structure have been carried out in different countries, predominantly in the U.S. and Canada (Keeler, 1974; Harris, 1977; Friedlander et al., 1981; Caves et al., 1981; Jara-Diaz et al., 1981; Braeutigam et al., 1984; Dodgson, 1985 and Keaton, 1991). They unequivocally point to advantages of large-scale production both in infrastructure and operating service. But it is impossible to make deductions from these studies as to the amount of scale advantages that would be lost in letting competition in the market guide operations on the Swedish railway network.

Therefore, to allow analysis a set of simple cost models are developed. Let cost of a train circuit on a line between two points be represented by the model

$$Y = A + BX, X \geq X_0 \qquad (1)$$

X is here number of wagons/coaches and X_0 maximum train size in number of wagons/coaches. For the sake of simplicity an average load per wagon and an average number of passengers per coach is assumed. A is furthermore a fixed incremental train cost and B a variable unit cost per wagon/coach. A includes capital and operating costs of locomotive per circuit (depreciation, energy for the locomotives' own propulsion, service and maintenance, infrastructure wear of the locomotive, driver, other staff independent of X) and the fixed incremental costs a train circuit causes in safety, signalling and control systems. B includes capital and operating costs per wagon/coach plus incremental costs per wagon/coach (depreciation, energy consumption, service and maintenance, handling of freight or passengers, wear on infrastructure and wagon control).

Let us now look at a duopoly situation. This is realistic considering that demand in Sweden is of a limited extent. No difficulties of principle to extend the model to an arbitrary number of competitors exist. Another condition is that SJ's competitor, considering the extent of capital needed and the considerable element of risk, sets up service within a limited area of the network, only. There, on the other hand, the establishment is assumed to be large enough to supply customers with complete railway service. A third condition is that the recently established competitor is supposed to use the same production technology as SJ and a fourth condition, finally, is that the two duopolists in essential dimensions produce the same quality of service.

Suppose now that SJ as a monopolist from the beginning supplies service within a certain product group on a certain line where demand corresponds to N trains per year and to X wagons/coaches per train and that cost per train circuit follows the function (1). With a fixed incremental line-specific product-group cost C, the yearly cost of this transport activity would be

$$Y_M = C + N(A + BX),\ X \leq X_0 \tag{2}$$

where X_0 is the maximum train size. Now suppose that the monopolistic situation turns into a competitive situation (duopoly) between two competing operators one of whom is SJ. Suppose furthermore that the same demand as in the initial stage is now shared between the competitors so that SJ's share corresponds to N_1 trains with on the average X_1 wagons/coaches per train and that of the other operator to N_2 trains with on the average X_2 wagons/coaches per train. Total cost of duopoly for the same effective production measured eg in tonnes or passenger-kilometres would then be analogously formulated as

$$Y_D = \sum_{k=1}^{2} Y_k = \sum_{k=1}^{2} \left[C_k + N_k\left(A_k + B_k X_k\right)\right], \qquad X \leq X_0 \tag{3}$$

where index k=1 is supposed to represent SJ and k=2 the competing enterprise having established itself on the line. Y_1 then is SJ's yearly cost and Y_2 that of the competitor.

A suitable measure for lost scale advantages from a transition to duopoly is the ratio

$$Y = \frac{Y_D}{Y_M} \tag{4}$$

This ratio may simultaneously be considered representing the minimum loss of large-scale production advantages from a transition to competition in the market since an analogous ratio in case of oligopoly would assume a greater value than (4). Society's primary motive for competition in the market is efficiency. With this type of competition advantages of large scale-production are lost. So, the question is how much competition must press down costs in order to compensate for these losses. A simple development of (4) demonstrates that the proportion in which costs must be reduced, here referred to as the compensatory potential of rationalisation, is equal to

$$P = 1 - \frac{1}{Y} \tag{5}$$

The measures Y and P may contribute to elucidate if establishment of competition in the market will increase cost efficiency of total production. The other main question being analysed in this section is the connection between advantages of large-scale production and potential establishers' propensity to enter the market. Under the given conditions a potential establisher's propensity of entering may be demonstrated by SJ's cost advantage defined as the ratio

$$K = \frac{\dfrac{Y_2}{N_2 X_2}}{\dfrac{Y_1}{N_1 X_1}} \tag{6}$$

ie the establisher's unit cost divided by that of SJ for given market shares. Particularly, at equal market shares (6) represents SJ's cost advantage through large-scale production advantages that emanate from network effects of other services and that the establisher apparently lacks.

Basic data for calculations have been provided by SJ's Staff Financing and Accounting according to our directions. Calculations have been carried out according to two different variants of sharing demand. Alternative one implies that if service today corresponds to n trains in a closed segment, SJ has in this segment got one train more than the competitor at uneven values of n, whereas the number of trains is equally shared at even values of n. Alternative two for sharing demand implies that present demand in each closed segment has been equally shared by competitors. This principle implies that if supply is n trains in one segment of present day service, each competitor will get n/2 trains if n is an even number and (n+1)/2 trains if n is uneven. Alternative one represents a cautious establishment strategy which gives SJ a bigger market share than the competitor and which results in a somewhat lower total cost of production whereas alternative two demonstrates a more aggressive establishment strategy which results in equal shares. Cost responsibility for the operator who is assumed to appear as a competitor to SJ in the duopolistic situation has been defined as follows:

- Traction and locomotive drivers, wagons/coaches and their staff, staff at stations, depot services, workshops for light current repairs, sales of tickets as well as some other miscellaneous items are treated as individual functions/resources. The enterprise itself is responsible for financing, investment and commercial risk. Competing enterprises are assumed to sell each other's tickets.
- Energy supply, fixed facilities at stations, traffic control, terminals, yards, workshops for heavy repairs and periodical overhaul, reservation system for seats as well as information systems are treated as common function/resource put a the enterprise's disposal at a unit price corresponding to SJ's cost price.

Four cases have been calculated: High-speed passenger service between Stockholm and Göteborg (case 1), Conventional long-distance service between Stockholm and Göteborg (case 2), Combined transport between Stockholm and Göteborg (case 3), Wagon-load transport between Hallsberg and Helsingborg (case 4). The results of the calculations are summarised in Table 1. In the table it may be seen that SJ's production advantage, also after an establisher being assumed to have captured half of the service exposed to competition, still is considerable even in the case which is the most favourable to the establisher, ie the aggressive establishment strategy (Alt. 2). In order to compensate this cost disadvantage the establisher must either take up other services in order to create cost reducing network effects or use newer and more cost-effective technique which can not be emulated by SJ. As it is not likely that there are enterprises which have these possibilities, it is not likely either that there are operators who will enter the market in the combinations of product and market segments where the calculation assumptions are valid.

Table 1 **Key factors of a cautious establishment strategy (Alt. 1) and an aggressive establishment strategy (Alt. 2) in four case studies**

Product	Lost economies of scale (Y)		Compensatory potential of rationalisation (P)		SJ's cost advantage (K)	
	A: Alt. 1	B: Alt. 2	C: Alt. 1	D: Alt. 2	E: Alt. 1	F: Alt. 2
Case 1	1.17	1.42	0.15	0.30	1.74	1.16
Case 2	1.21	1.24	0.17	0.19	1.35	1.21
Case 3	1.17	1.52	0.15	0.34	1.29	1.25
Case 4		1.36		0.26		1.17

It may furthermore be inferred that an introduction of duopoly would cause an important loss of economies of scale (Columns A and B). In order for competition in the market to improve total efficiency, compensatory cost rationalisations (Columns C and D) must be made, which seem difficult to realise under the circumstances given by the assumptions for the calculations, since competitive pressure is important in the majority of combinations of product and market segments. This is particularly true in the case of aggressive establishment strategy which seems to be the strategy that an establisher may be expected to choose as it implies the least unfavourable cost disadvantage (Columns E and F).

The conclusions drawn here seem valid for all products but domestic combined transport. In this market there are large volumes of freight which by a transition to competition in the market may be transferred to combined service both from regular long-distance road transport and from wagon-load service. These additional volumes may imply that competition in the market can be established here without any large-scale production advantages being lost. Besides, combined transport may have an extra potential for cost rationalisations by technological development which can be stimulated by i. a. competition.

Other costs

Internal competition cannot be introduced without significant transaction costs for reducing barriers to competition, transaction costs for controlling competition and costs of duplication of functions and lower of resources. The strategy of creating perfect conditions for internal competition on the railway network by separating infrastructure and traffic must, in order to succeed, be able to reduce barriers to establishment and to control competition. The interferences in the railway organisation caused by this incur, however, that a considerable mass of resources and costs can no longer be exposed to competitive pressure depending on its being moved from the part of the vertical system that competes for demand in the freight transport and travel markets to a part of the system which is not exposed to competition. In other words: internal competition in one part of the vertical system of transport production can be bought at the price of an eliminated competitive pressure in another. Transfer of resources from the part of the system exposed to competition to the part not exposed to competition can with good reasons be expected to imply loss of efficiency in the mass of resources that is moved. These losses are another component in the costs of internal competition.

Based on data from The Ministry of Transport and Communications, it is possible to show that in a perfect situation for internal competition only about 25 per cent of SJ's present total costs would be exposed to external (and internal) competition. If BV and SJ are seen as a vertical, but not integrated, production system, only about 17 per cent of the total costs would be exposed to competition. Less far-reaching, and thereby less perfect, reorganisation of transport production may possibly increase the percentage of costs to be exposed to competition. It seems, however, hard to believe that this part could amount to more than about 50 per cent of SJ's present costs.

Considering jointly external competition, cost scale advantages and other costs, only three segments are expected to improve efficiency from opening the market for internal competition: (1) Domestic combined freight transport (competition in the market), (2) feeder service in wagon-load freight transport (competitive tendering) and (3) unprofitable passenger services bought by society (competitive tendering).

WHY VERTICAL INTEGRATION?

Introduction

The two preceding parts have raised several questions as to the possibilities and effects of deregulating railway transport. This leads to the following questions: Was the division of the railway in 1989 justified? Are there alternatives to the 'Swedish railway model' which better fulfil the requirement of efficiency? The fundamental question to be raised is why railways never ever were operated the way it is now and will be in the future in Sweden. Traditionally, railways have been by vertical integration between track and operations, ie infrastructure and transport service being produced by the same enterprise. The railway demonstrates in this respect an industrial organisation that is different from other forms of traffic. Within air and sea traffic, infrastructure has traditionally been separated from production of transport services and not only that; it is not unusual that owners of ports and airports act only as landlords, and leave most of the operations to other actors/operators. This has apparently had its economic advantages. A principal question is, in other words: if the competitive model in reality can offer the advantages propounded by its supporters, then why has it not emerged before, not even in those parts of the world and during periods in which public regulations would not have interfered with its introduction? An important part of the answer is related to the driving forces that lead to vertical integration within different forms of production activities.

Driving forces of vertical integration

Even if the ideas behind the Swedish railway model have never been properly accounted for, they seem to build on the following assumptions. Firstly, the provider of infrastructure acts rationally and strives at minimising costs of production. Secondly, the cost structure of this provider is characterised by linearity with respect to transport volume. Thirdly, the operator's cost structure is also characterised by linearity and divisibility in the same way as it is often assumed for road transport. Fourthly, the public costs for executing authority functions, 'organising' the market on the tracks and controlling that 'competitive neutrality' is respected, are not too high.

Given those circumstances, traditional theory indicates that no economic benefits may be achieved by vertical integration between infrastructure and operations. Nor can there be any private economic gains as a result of such an integration. It does not matter whether the provider of infrastructure acts as a monopolist or in a market exposed to competition; the theory indicates the same result (Shughart, 1991). Thus, a theoretical basis for the Swedish model exists. Besides, as the provider of infrastructure is a public authority, it is also—in principle—possible to assert that one can prevent it from acting as a monopolist, eg by stipulating that prices for utilisation of track be equal to marginal costs.

However, there are a number snags in this argument. Several of the assumptions above are in reality not satisfied. The fact is that many of the traditional factors that economic theory indicates as powerful driving forces behind vertical integration in all likelihood are also present within the railway. These driving forces ultimately arise from the possibilities to reduce total cost of production. This also implies that vertical integration under those conditions leads to lower costs thereby being compatible with the principal objective of economic efficiency.

The following presentation builds on the assumption that both the upstream operator (provider of infrastructure) and the downstream operator/operators are maximising profits. Six factors are of primary relevance (Perry, 1989, Shughart, 1991):

- Substitutability between resource inputs into track and operations for producing a given volume;
- Scale economies at the operator level;
- External effects (congestion) at the operator level;
- Uncertainty and the costs associated therewith;
- Exploiting possibilities of price discrimination;
- Transaction costs.

Substitutability between factors of production

The first reason for vertical integration is that costs of transport operations are not independent of those of the infrastructure service. This means that there is a need for close co-ordination of resource inputs into track and operations in order to minimise total costs. This is true in the short, medium and long run. In the short run there is the problem of agreeing on periods and extent of track maintenance. In the medium-term perspective, growing transport volumes will sooner or later make it cost-efficient to invest in more capacity, eg by laying another passing track, additional tracks or by upgrading the signalling system. Reduced need for rolling stock, etc. means savings in operations which may well compensate the investment costs. The same situation is at hand in the long run. An increase in the maximum axle load eg means substantial savings in transport production cost. It is unlikely that an autonomous infrastructure provider will take these effects into account, ultimately as he is a monopolist, and therefore is not concerned with the overall minimisation of costs in the railway.

Scale economies at operator level

Driving forces towards vertical integration are accentuated if there is only one operator, eg as a result of decreasing unit costs in transport production. The non-linear cost structure at the operator level increases the need for co-ordination in order to produce at lowest possible costs, and, besides, the existence of only one operator facilitates that. As pointed out above, indications are that the operators' activities are by economies of scale. There will, consequently, not be any real competition on the track. There will be either only one operator or several operators, who will act as local monopolists on separate parts of the track.

Congestion

It is a well-known problem that high utilisation of roads and runways creates congestion so that different operators' costs become interdependent. The existence of these external effects often implies that infrastructure is inefficiently used. The classical solution to this is to introduce congestion charges. Another solution would be to integrate operations and infrastructure. Both solutions lead to lower costs and are therefore economically justified. Both solutions are also equivalent from a distributional point of view if the congestion charges are paid to the provider of infrastructure. Bearing in mind that the congestion effects are relatively seen much more important within the railway one would consider the latter solution more relevant to this kind of traffic.

Uncertainty

The fourth driving force of vertical integration is uncertainty. Uncertainty costs seem to be exceptionally high within the railway due to the long technical life of investments and the high proportion of irreversible costs. The infrastructure provider is much more affected by this problem than the operators. The infrastructure provider's problem is accentuated by being much further away from the market than the operators—it has no direct contact with the customers—and by its fixed assets having much longer lives than those of the operators. On the other hand, the operators are very dependent on the infrastructure provider's planning of and decisions on the provision of additional track capacity and is also affected by the fact that planning and implementing capacity-increasing investments take a long time. A vertical integration between infrastructure and operations implies under these circumstances that some risks may be eliminated and that total cost of uncertainty may be reduced. Again, vertical integration is for this reason economically justified.

Price discrimination

The fifth reason for vertical integration is actually closely connected to the uncertainty problem and to the quest for reducing the cost thereof. Price discrimination does not only refer to the situation where different transport buyers are charged different prices for the same service, but also to indirect price discrimination through the exploitation of land and real-estate adjacent to railway premises. Operators may due to competitive pressure from other means of transport and alternatives be expected to be dependent on price discrimination as a way to increase revenues. The operators' possibilities to do so increase the better adapted infrastructure is to market demands and vice versa. There is therefore a natural driving force for co-ordinating transport and infrastructure planning in order to maximise the opportunities for price discrimination.

Transaction costs

Transaction costs refer to the costs of 'maintaining and running' a market. They include costs of following the development of prices, purchasing, contracting, conflict resolution, etc. If transaction costs are very high it may be rational to solve the production problem by circumventing the market, eg by vertical integration. In the Swedish model transaction costs arise in two areas. The first area was dealt with earlier, regarding relations between different operators. That cost component is the more important part and most likely to be significant. The other area regards relations between the operator/operators and the provider of infrastructure and embraces i.a. costs of negotiating and implementing agreements on track maintenance and track quality, eg in the form of agreements on track allocation which are now concluded between BV and SJ. But this component also includes all the other activities that different interested parties—including SJ, municipalities, politicians, industry—engage in to influence the quality and capacity of the railway network. The extensive lobbying—ie directly unproductive profit-seeking activities—which the Swedish model has indeed provoked, is in other words, also an example of a transaction cost.

Relevance to the Swedish railway model

In the presentation above it has been assumed that there are a number of operators and a provider of infrastructure. All these actors are supposed to maximise profits, and the aim has been to point, at the driving forces which make the operators and the provider of infrastructure strive to integrate their activities and at this being compatible with the objective of economic efficiency.

Now the question is what relevance this has to the Swedish model. It involves a provider of infrastructure who does not act as a monopolist maximising profit but who is 'rational'. It is conceivable that, if it were possible to make a publicly-owned monopoly act rationally, many of the driving forces identified above would be reduced or even eliminated. Under favourable conditions this is true in all the above-mentioned forces behind vertical integration, except for transaction costs. Given those circumstances, maybe even the Swedish model would be rational. Again, the problem is the realism of the assumption that has to be made. Thus, experience indicates that one can not expect a public monopoly to act rationally in the way it must be assumed. Now, if it were assumed that a public monopoly would be able to act rationally, then why is it considered desirable to introduce competition on the track?

Finally it is worth mentioning that the discussion above may give a part of the explanation why vertical integration is not of the same importance for air, sea and road traffic. Transaction costs are considerable lower than in railway service, due to the fact that external effects arising in connection with bottlenecks are not at all of the same chronic nature. Possible substitution between resources being used in producing transport and infrastructure services is of course not negligible for other kinds of traffic but very likely of less importance. Furthermore, other producers of transport services are not in the same way as the railway dependent on the provider of infrastructure and vice versa, so uncertainty costs are of a smaller dimension. This in turn is related to the condition that the cost structure of operators in other transport sectors are much closer to the ideal linear and divisible form. However, the relative proportion of infrastructure costs is ultimately the decisive difference between the railway and other kinds of traffic. Within the railway, costs directly connected to tracks, catenaries, telephone lines, stations, control offices, etc. amount to almost 50 per cent of total costs; in other kinds of traffic the corresponding proportion is only about 5-10 per cent.

CONCLUDING WORDS

The transport policy that has been implemented is ultimately based on the assumption that the railway is comparable to other modes. That is a mistake. The railway's infrastructure and that of other modes are in essence different in that railway vehicles have only one degree of freedom of movement whereas other types of vehicles have several degrees of freedom.

The deficiency of the Swedish model will harm the railway in the short run as well as in the long-term perspective. In the short run it is the proposal to introduce competition on the track that represents the threat. The proposed deregulation will, in principle, lead to a separation of decisions on train plans and of traffic control from the control of transport production. Since the service supplied and demanded—namely infrastructure capacity—in reality can not be exactly specified, traffic planning and traffic control will be characterised by arbitrariness and uncertainty from the operator's point of view. This means that, regardless of whether competition on the track will materialise or not, which is doubtful, uncertainty will increase. From this follows that the transport enterprise will diminish investments and raise prices.

In the long run, it is the separation between infrastructure and train operations that is detrimental. This division leads to a distorted direction of investments, a situation that will be aggravated by the investments being made on the basis of socio-economic benefit-cost analysis and regardless of the need for covering infrastructure costs. A number of investments will be made that should never have seen the light of the day whereas other investments of great importance to the railway's future competitive situation run the risk of not being implemented. The importance of this issue is connected to the infrastructure's comparatively high share of total costs of producing railway service, in round figures about 50 per cent.

In summary, the conclusion is that Sweden chose the wrong way of reforming transport policy in 1988. The commercialisation approach should have been chosen instead of the vertical disintegration approach. The former means, that the railway would be allowed to operate in the form of a vertically integrated enterprise but on commercial conditions and in free competition with other modes.

The commercialisation approach does not mean that only one enterprise will work within the railway sector. On the contrary, this approach may lead to a situation where several different companies are involved in the production of railway transport services, and it may even happen that several operators will use the same track. However, an important difference is that the different companies will co-operate rather than compete with each other. The fundamental condition is only that decisions are based on sound commercial principles and not on political considerations. Competition offered by other modes and alternatives, ie external competition, is the main barrier to this co-operation becoming detrimental to the shipper.

ACKNOWLEDGMENTS

This paper is based on a book that was first published in Swedish in 1993 by SNS Förlag, Stockholm. The book was later translated into English and published under the title: "Swedish Rail Policy—A Critical Study" in Meddelande 78 from the Department of Transportation and Logistics, Chalmers University of Technology. Both the original research effort and the translation was made possible through grants from the Swedish association FJT (Framtida Järnvägstrafik = Future Rail Traffic). The authors want to thank FJT and SNS for their support and encouragement.

REFERENCES

Aaker, D.E. (1988) *Strategic Market Management,* New York, N.Y.: Wiley.

Abell, D.E. (1980) *Defining the Business,* Englewood Cliffs, NJ: Prentice-Hall.

Brauetigam, R.R., Daughety, A.F. and Turnquist, M.A. (1984) "A Firm Specific Analysis of Economics of Density in the US Railroad Industry", *Journal of Industrial Economics,* vol. 23, ss 2-20.

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

Dodgson, J.S. (1985) "A Survey of Recent Developments in the Measurement of Rail Total Factor Productivity", in Button, K.J. & Pitfield, D.E. (Eds), *Internat. Railway Economics,* Aldershot: Gower.

Friedlaender, A.F. and Spady, R.H. (1981) *"Freight Transport Regulation",* Cambr., Mass.: MIT Press.

Harris, R. (1977) "Economies of Traffic Density in the Rail Freight Industry", *Bell Journal of Economics,* Vol. 8, ss 556-63.

Jara-Diaz, S. and Winston, C. (1981) "Multiproduct Transportation Cost Functions: Scale and Scope in Railway Operations", i Blattner, N. et al. (red.), *Eighth European Association for Research in Industrial Economics,* Vol. 1, Basel.

Keaton, M.H. (1991) "Economies of Density and Service Levels on US Railroads: An Experimental Analysis", *Logistics and Transportation Review,* Vol. 26, ss 211-27.

Keeler, T.E. (1974) "Railroad Costs, Return to Scale and Excess Capacity", *Review of Economic Statistics,* Vol. 56, ss 201-08.

Leibenstein, H. (1966) "Allocative Efficiency v.s. X-Efficiency", *Am. Ec. Review,* Vol. 56, ss 392-415.

Leibenstein, H. (1976) *Beyond Economic Man,* Cambridge, Mass.: Harvard University Press.

Nash, C.A. (1990) "Role of Rail in Future Transport Policy", *Rees Jeffreys Discussion Paper no. 12,* Transport Studies Unit, University of Oxford.

Nash, C.A. (1991) *"Rail Policy in the European Community"*, Working Paper 324, Institute for Transport Studies, University of Leeds.

Perry, M.K. (1989) "Vertical Integration: Determinants and Effects", Chapter 4 in Schmalensee, R. and Willig, R.D., *Handbook of Industrial Organization, Volume I,* Amsterdam, Elsevier Sci. Publ. B.V.

Porter, M.E. (1980) *Competitive Strategy: Techniques for Analyzing Industries and Competitors,* New York, N.Y.: The Free Press.

Shepherd, W.G. (1979) *The Economics of Industrial Organization,* Englewood Cliffs, NJ: Prentice-Hall.

Shughart II, W.F. (1991) *The Organization of Industry,* Homewood, I1.: BPI/Irwin.